



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 116820

TO: Vanessa L Ford
Location: REM-3B25&3C18
Art Unit: 1645
Thursday, March 18, 2004

Case Serial Number: 09/543407

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen E01A69
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

Search Notes

RUSH

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

116820

From: Page, Thurman
Sent: Friday, March 12, 2004 2:35 PM
To: STIC-Biotech/ChemLib
Cc: Ford, Vanessa
Subject: FW: In re: 09/543407 sequence search

Importance: High

CRFE

Rush search approved.

Vanessa,
please resend the reason for rush to me for my record.
thanks

Thurman K. Page
SPE Art Units 1615 & 1616
Technology Center 1600

-----Original Message-----

From: Ford, Vanessa
Sent: Friday, March 12, 2004 1:52 PM
T : Page, Thurman
Subject: FW: In re: 09/543407 sequence search

Christina is out of the Office today.

-----Original Message-----

From: Ford, Vanessa
Sent: Friday, March 12, 2004 1:50 PM
To: Chan, Christina
Subject: In re: 09/543407 sequence search

Please search: SEQ ID NOs: 1, 2, 3, 4, 9, 11, 13, 15, 17, 19. please include interference searches. Please Rush.

Biotechnology Patent Examiner
Office: REM 3B25
Mailbox: REM 3C18
Phone: 571.272.0857
Art Unit: 1645

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgac.....ccacgcctaacagctactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vr:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_mus:*
41: em_hcg_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	144909	144909 Sequence 58
2	456	100.0	2067	1 SEU43280	U43280 Salmone11
3	448	98.2	5103	1 STAJ2301	AJ002301 Salmone11
4	448	98.2	22411	1 AE008749	AE008749 Salmone11
5	446.4	97.9	254050	1 AL627269	AL627269 Salmone11
6	446.4	97.9	301983	1 AE016840	AE016840 Salmone11
7	440.6	96.6	1048	1 STAF6A	AU000514 Salmone11
8	431.8	72.8	361	6 144908	144908 Sequence 56
9	310.4	68.1	2889	1 CSP515700	AJ515700 Citrobact
10	262.8	57.6	2920	1 CFS15701	AJ515701 Citrobact
11	256	56.1	4680	1 ECCSGABDG	X60754 E.coli csgG
12	256	56.1	10346	1 AE000205	AE000205 Escherich
13	256	56.1	15047	1 D90741	D90741 Escherichia
14	254.4	55.8	456	6 AX814811	AX814811 Sequence
15	254.4	55.8	648	1 ECOC8GAA	L04979 Escherichia
16	247.8	54.3	1711	1 AF275733	AF275733 Escherich
17	247.8	54.3	10190	1 AE005315	AE005315 Escherich
18	247.8	54.3	306358	1 AE016759	AE016759 Escherich
19	247.8	54.3	327773	1 AP002554	AP002554 Escherich
20	229	50.2	230	1 SEU53207	U53207 Salmone11
21	217.6	47.7	2883	1 ESAS15702	AJ515702 Enterobac
22	190.4	41.8	10370	1 AE015131	AE015131 Shigella
23	190.4	41.8	292504	1 AE016981	AE016981 Shigella
24	167.4	36.7	437	1 AF237726	AF237726 Shigella
25	93.2	20.4	19201	1 D90742	D90742 Escherichia
26	64.6	14.2	1212	1 ECO131756	AJ131756 Escherich
27	48.2	10.6	78	6 AX814809	AX814809 Sequence
28	44.8	9.8	72	6 AX814798	AX814798 Sequence
29	44.4	9.6	2000	6 AX655393	AX655393 Sequence
30	36.2	7.9	1425	1 AF343445	AF343445 Lactobaci
31	36.2	7.9	302156	3 AC116977	AC116977 Dictyoste
32	36	7.9	168267	2 AC146009	AC146009 Pan trogl
33	35.6	7.6	301214	1 AE016786	AE016786 Pseudomon
34	34.6	7.6	11160	1 AE008209	AE008209 Agrobacte
35	34.6	7.6	11476	1 AE009405	AE009405 Agrobacte
36	34.4	7.5	151828	2 AC146008	AC146008 Pan trogl
37	34.2	7.5	1458	1 AF173043	AF173043 Oryz sat
38	34	7.5	956	8 AK058943	AK058943 Oryza sat
39	34	7.5	1344	6 BD243128	BD243128 Hyperdens
40	34	7.5	1344	6 BD243666	BD243666 Hyperdens
41	34	7.5	1344	6 AR403552	AR403552 Sequence
42	34	7.5	1344	6 AX287002	AX287002 Sequence
43	34	7.5	1344	6 AX343147	AX343147 Sequence
44	34	7.5	1344	6 AX370155	AX370155 Sequence
45	34	7.5	1344	6 AX536272	AX536272 Sequence

ALIGNMENTS

RESULT 1
LOCUS 144909
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION 144909
VERSION 144909.1 GI:2469622
KEYWORDS
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
Doran,J.L., Kay,W.W., Collins,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
Salmone11
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

FEATURES
source
Location/Qualifiers
1..456
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 456; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-120;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGCGAGCATTCGACGACATCTAGTCTGGCAGTGGCT 60
Db 1 ATGAACCTTTTAAAGTGCGAGCATTCGACGACATCTAGTCTGGCAGTGGCT 60
QY 61 GGGCTGCTCCACATAGGGGCGGCGGTAATCATTAACGGCGGCAATAGTTCCGCG 120
Db 61 GGGCTGCTCCACATAGGGGCGGCGGTAATCATTAACGGCGGCAATAGTTCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTTATAGTACGTTCCGCTAACGCTGCTTCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTTATAGTACGTTCCGCTAACGCTGCTTCTGCA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGGCCGAT 240
Db 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGGCCGAT 240
QY 241 GTAGGCGAGGCTCGGATTAATAGTACTTAACTGATCTGCGCAATGCTTCAAAATAT 300
Db 241 GTAGGCGAGGCTCGGATTAATAGTACTTAACTGATCTGCGCAATGCTTCAAAATAT 300
QY 301 GCCACATGACCATGAGTAACGCTTAACTCCGATATTCTGTCGCAATAGCGCGGT 360
Db 301 GCCACATGACCATGAGTAACGCTTAACTCCGATATTCTGTCGCAATAGCGCGGT 360
QY 361 AATTAACGCGCGCTGTTATCAGACCGCATCTGATTCAGCGTATGCTGCTGAGTT 420
Db 361 AATTAACGCGCGCTGTTATCAGACCGCATCTGATTCAGCGTATGCTGCTGAGTT 420
QY 421 GGTITTTGGCAACACGCCACGGCTAACCGATTTAA 456
Db 421 GGTITTTGGCAACACGCCACGGCTAACCGATTTAA 456

RESULT 2
SEU43280 2067 bp DNA linear BCT 14-FEB-1996
LOCUS
DEFINITION
Salmonella enteritidis agfBAC operon: fimbria-like protein precursor (agfB), thin aggregative fimbriae precursor (agfA), and AgfC (agfC) genes, complete cds.
ACCESSION
U43280.1 GI:1184712
VERSION
KEYWORDS
SOURCE
ORGANISM
Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS
Doran, J.L., Collinson, S.K., Burian, J., Sarios, G., Todd, E.C., Munro, C.K., Kay, C.W., Baner, P.A., Peterkin, P.I., and Kay, W.W.
TITLE
DNA-based diagnostic tests for Salmonella species targeting agfA, the structural gene for thin, aggregative fimbriae
JOURNAL
J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
MEDLINE
94013373
PUBMED
8104955
REFERENCE
2 (bases 1 to 2067)
AUTHORS
Collinson, S.K., Cloutier, S.C., Doran, J.L., Baner, P.A. and Kay, W.W.
TITLE
Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae
JOURNAL
J. Bacteriol. 178 (3), 662-667 (1996)
MEDLINE
96146512
PUBMED
8550497
REFERENCE
3 (bases 1 to 2067)

AUTHORS
Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
TITLE
Direct Submission
JOURNAL
Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg., Victoria, BC V8W 3P6, Canada
FEATURES
source
Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrc"
571..576
598..603
696..1151
/gene="agfB"
/note="fimbria-like protein precursor"
/codon_start=1
/transl_table=11
/product="AgfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKKKLLFPMVLTILGPGIATATYDILARSEYNPAVELSSSEN QALIGOVGTDSARVROGSKLTSVTSOBGANNRAKDYDGNTPATIEOTGNANDA SISQAYGNSAIIQKSGNKNANTIOYGTOKTAVVQKSHMALRVQR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
/product="AgfB"
/note="putative signal sequence of 21 amino acids; fimbria-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbria subunit precursor; major fimbrial subunit of thin aggregative fimbriae precursor"
/codon_start=1
/transl_table=11
/product="AgfA fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLLKVAAPFAIVSGSALAGVVPOMGGGNGGNSGSPDST LSIYQGSANAALALOSDARKSETTIGSGYGDADVGCGADNSTIETLNGRRNAT IDQWNAKNSDIYTGQYGSNNALVNGTASDSSVWROVGFNNATANQY"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/product="AgfA fimbria"
/note="thin aggregative fimbria subunit; major fimbrial subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=11
/product="AgfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MTHTLLAALASNOITFTTQGDITVYIPVTLNBPVCQVOIT SVYDVGQSHTOQKOTLSLPANQPIELSRISVNTSSSDSKITVYDSGSLHSQ MPESAQ"

ORIGIN

Query Match 100.0%; Score 456; DB 1; Length 2067;
Best Local Similarity 100.0%; Pred. No. 1.8e-120;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
Qy	1	ATGAACTTTTAAAGTGGCGATTCGACGATTCGAGTTTCTGGCAGCTCTGCT	60
Db	1193	ATGAACTTTTAAAGTGGCGATTCGACGATTCGAGTTTCTGGCAGCTCTGCT	1252
Qy	61	GGCGTCGTTCCAGCAATGGGGGGGGGTAATCATTAACGGCGCGGCAATGTTCCGGC	120
Db	1253	GGCGTCGTTCCAGCAATGGGGGGGGGTAATCATTAACGGCGCGGCAATGTTCCGGC	1312
Qy	121	CCGAGCTCAACGTTGAGCATTATCAGTACCGTTCCGCTAACGCTGCTCTGCA	180
Db	1313	CCGAGCTCAACGTTGAGCATTATCAGTACCGTTCCGCTAACGCTGCTCTGCA	1372
Qy	181	AGCGATGCGCTTAATCTGAAAACGACCTTACCAAGCGGTTATGTTAACGGCGCGAT	240
Db	1373	AGCGATGCGCTTAATCTGAAAACGACCTTACCAAGCGGTTATGTTAACGGCGCGAT	1432
Qy	241	GTAGCGCAGGGTGGGATTAATGACTATTAAGTCACTGAGTCAAGATGTTTCAAAATAT	300
Db	1433	GTAGCGCAGGGTGGGATTAATGACTATTAAGTCACTGAGTCAAGATGTTTCAAAATAT	1492
Qy	301	GCCACATCGACGACGTGGAACGCTAAACTCCGATATTACTGTCGGCCAAATACGGCGGT	360
Db	1493	GCCACATCGACGACGTGGAACGCTAAACTCCGATATTACTGTCGGCCAAATACGGCGGT	1552
Qy	361	AATACGCGCGCTGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGTCAGGTT	420
Db	1553	AATACGCGCGCTGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGTCAGGTT	1612
Qy	421	GGTTTGGCAACAGCGCAGCGGCTAACGATTA	456
Db	1613	GGTTTGGCAACAGCGCAGCGGCTAACGATTA	1648
RESULT 3	STA2301	5103 bp DNA linear BCT 15-NOV-2000	
LOCUS	Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and k		
DEFINITION	csgG genes.		
ACCESSION	AJ002301		
VERSION	AJ002301.1 GI:2739232		
KEYWORDS	csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene; csgG gene.		
SOURCE	Salmonella typhimurium		
ORGANISM	Salmonella typhimurium		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.		
AUTHORS	Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Normark, S.		
TITLE	Curl1 fibers are highly conserved between Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation		
JOURNAL	J. Bacteriol. 180 (3), 722-731 (1998)		
MEDLINE	98117058		
PUBMED	9457880		
REFERENCE	2 (bases 1 to 5103)		
AUTHORS	Romling, U.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-OCT-1997) Romling U., Department of Bacteriology, Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN		
FEATURES	location/Qualifiers		
SOURCE	1..5103		
	/organism="Salmonella typhimurium"		
	/mol_type="genomic DNA"		
	/strain="SR-11"		
	/db_xref="taxon:602"		
	<1..332		
	/note="ORF179a"		
	/codon_start=2		
	/transl_table=1		
	/product="hypothetical protein"		
	/protein_id="CAA05311.1"		
CDS			
	/db_xref="GI:2739233"		
	/db_xref="GOA:O54290"		
	/db_xref="SWISS-PROT:O54290"		
	/translation="CTLSCMVVALVVIHGSALAPHLIEIVSYLTGIVAFIMCIQAKO LVLSFVGTPIGACATFAGOGDWRLVPLSLAGIIFGAMKNSGLMLASRREHSANT AVRK"		
gene	complement (361.. 1194)		
CDS	/gene="csgG"		
	complement (361.. 1194)		
	/gene="csgG"		
	/function="lipoprotein"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=1		
	/product="CsgG protein"		
	/protein_id="CAA05312.1"		
	/db_xref="GI:2739234"		
	/db_xref="GOA:O54291"		
	/db_xref="SWISS-PROT:O54291"		
	/translation="MPRLILVAILISGLTAPKQAAKPTLMPRAQSYDILTHLP PTGKIFVSVNIODETGFQKPYPSANSTAVPSATMLTALDQSRFTLEKQGLQ NLINERKIIRAQENGTVAMNRIPLQSLTAANIMVGSIIIGSVNKSQGVARFVG IGADTQYQLDQIIVNLRVNVSTGEILSVTSKTIISYEQACVFFPIIDYQRLLEG IGTYSNBPMLCLMSAIEYGVIFLINDIGRLMDLONKADRONDIIVKYEELSVPE S"		
gene	complement (1221.. 1637)		
CDS	/gene="csgF"		
	complement (1221.. 1637)		
	/gene="csgF"		
	/codon_start=1		
	/transl_table=1		
	/product="CsgF protein"		
	/protein_id="CAA05313.1"		
	/db_xref="GI:2739235"		
	/db_xref="SWISS-PROT:O54292"		
	/translation="MRVKAHVLIMLFSPLTWAGMTFFQNPFGGNPNNGSEFLNS AOAONSKDPAYNDPFIETPSPALDNFTQAIQSOILGLGLTNITVGRRVNTDFII DIANRDQLOLANTDRKTRGRTLEVSGLOTQSIDP"		
gene	complement (1664.. 2059)		
CDS	/gene="csgE"		
	complement (1664.. 2059)		
	/gene="csgE"		
	/codon_start=1		
	/transl_table=1		
	/product="CsgE protein"		
	/protein_id="CAA05314.1"		
	/db_xref="GI:2739236"		
	/db_xref="SWISS-PROT:O54293"		
	/translation="MKRYLTIWVAIELLPATGNLHANEVEVEVGLLTDHTVSSIGHE FYRASFQKMESEYTGNTITNERPARSGSVITITVNDVIFQFTLPFMKDFEKTIVF ALMOTERALKRRQIDOTLSTDLARDF"		
gene	complement (2064.. 2714)		
CDS	/gene="csgD"		
	complement (2064.. 2714)		
	/gene="csgD"		
	/function="transcriptional regulator"		
	/codon_start=1		
	/transl_table=1		
	/product="CsgD protein"		
	/protein_id="CAA05315.1"		
	/db_xref="GI:2739237"		
	/db_xref="GOA:O54294"		
	/db_xref="SWISS-PROT:O54294"		
	/translation="WFNEVSHSHHTLLLTTPKSLQATALLQHLKQSLATTKGLANIO RSLSDISAGCIVLMDMWDADKLIHYQDNLGRNNNIKTLTLNTPDYPVREIENNP HSNGEVPATDEOHVVSGLQIGLGEYFQOKLASVYITSGNVRVNSTSALITRHE KEILNKRIQASNNIEIARSLPISENTVKTLVNLFKIAYKNRQAVSWANDLNR"		
gene	complement (2869.. 2896)		
-10_signal	/gene="csgDERG"		
-35_signal	complement (2869.. 2874)		
	/gene="csgDERG"		
	complement (2891.. 2896)		
	/gene="csgDERG"		

```

gene          3344. .3376
              /gene="csGA"
-35_signal    3344. .3349
              /gene="csGBA"
-10_signal    3371. .3376
              /gene="csGBA"
gene          3469. .3924
              /gene="csGB"
CDS           3469. .3924
              /gene="csGB"
              /function="nucleator"
              /codon_start=1
              /evidence=experimental
              /transl_table=11
              /product="CsGB protein"
              /protein_id="CA05316.1"
              /db_xref="GI:2739238"
              /db_xref="GOA:P55226"
              /db_xref="SWISS-PROT:P55226"
              /translation="MKNLPLMPLTIIAGGIGIATAINVDLARSEYNPAVNEI.SKSRFR
              QALITQVDTNSARVROEGSKLSVISGCGNNRAKVDACNVAIPIETQGNANDA
              S3QSYGNSAIIIOKGSKNANITGYGKTAVVOKOSHAIIVTOR"
sig_peptide   3469. .3531
              /gene="csGB"
mat_peptide   3532. .3921
              /gene="csGB"
              /product="CsGB protein"
              /function="nucleator"
              /gene="csGA"
              /gene="csGA"
              /function="fiber subunit"
              /codon_start=1
              /evidence=experimental
              /transl_table=11
              /product="CsGA protein"
              /protein_id="CA05317.1"
              /db_xref="GI:2739239"
              /db_xref="GOA:P55225"
              /db_xref="SWISS-PROT:P55225"
              /translation="MKLKVAAFAAIVVSGSALAGVYPONGGNGNGGNSGPDST
              LSIYVGSNAALAILOSDARKSETITTSIGTGNADVGQADNSTIETLTONFRNNAT
              IDOWNKNSGITTYGQYGNNAALVNQITASDSSVMRVGNNATANYQ"
sig_peptide   3966. .4025
              /gene="csGA"
mat_peptide   4026. .4418
              /gene="csGA"
              /product="CsGA protein"
              /function="fiber subunit"
              /gene="csGB"
              /gene="csGB"
              /codon_start=1
              /transl_table=11
              /product="CsGB protein"
              /protein_id="CA05318.1"
              /db_xref="GI:2739240"
              /db_xref="SWISS-PROT:P55227"
              /translation="MHTLIIAASNSQITPTTTQGDIVYIIPVTLNPEPCVOVQL
              SVRDVGQSHITQOKQITLSIPANQFIETLSRLSVNISSESVKIIIVTSDGQSHTLSQO
              WPSNQ"
sig_peptide   4483. .4533
              /gene="csGC"
mat_peptide   4534. .4806
              /gene="csGC"
              /product="CsGC"
              /gene="ORF103"
              /gene="ORF103"
              /gene="ORF103"
              /codon_start=1

```

Query Match	Best Local Similarity	98.2%; Score 448; DB 1; Length 5103;
Matches 451; Conservative 0; Mismatches 5; Indels 0; Gaps 0;		
QY 1	ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTATGTTCTGGCAGTCTCGCT	60
Db 3966	ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTATGTTCTGGCAGTCTCGCT	4025
QY 61	GGCGTCGTTCCACATGAGGCGCGCGGCTATCTATAACGCGCGGCAATAGTTCGCG	120
Db 4026	GGCGTCGTTCCACATGAGGCGCGCGGCTATCTATAACGCGCGGCAATAGTTCGCG	4085
QY 121	CCGAGCTCAACGTTGAGCATTTATCAGTACGGTTCGCTACGCTGCGCTGCTCGCA	180
Db 4086	CCGAGTTTCAACGTTGAGCATTTATCAGTACGGTTCGCTACGCTGCGCTGCTCGCA	4145
QY 181	AGCGATGCCGTAATCTGAAACGACCATTAACCGACGCGGTTATGGTAAACGCGCGAT	240
Db 4146	AGCGATGCCGTAATCTGAAACGACCATTAACCGACGCGGTTATGGTAAACGCGCGAT	4205
QY 241	GTAAGCCAGGCGCGGTAATAGTATCTATTGAACTGACATGATGTTTACAAATAT	300
Db 4206	GTAAGCCAGGCGCGGTAATAGTATCTATTGAACTGACATGATGTTTACAAATAT	4265
QY 301	GCACCATTCGACAGTGTGAAACGCTAAACATCCGAGATTAATCTGCGCCAAATACGCGGT	360
Db 4266	GCACCATTCGACAGTGTGAAACGCTAAACATCCGAGATTAATCTGCGCCAAATACGCGGT	4325
QY 361	AATAACGCGCGGCTGTTATATCAGACCGCATCTGATTCAGCGTATAGTGTGCGT	420
Db 4326	AATAACGCGCGGCTGTTATATCAGACCGCATCTGATTCAGCGTATAGTGTGCGT	4385
QY 421	GGTTTGGCAACAAACGCAACGCGTAAACAGTATTA	456
Db 4386	GGTTTGGCAACAAACGCAACGCGTAAACAGTATTA	4421

RESULT 4

AE008749

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

CONSTRM

TITLE

JOURNAL

22411 bp

DNA

linear

BCT 23-Apr-2003

Salmonella typhimurium LT2, section 53 of 220 of the complete genome.

AE008749 AE006468

AE008749.1 GI:16419641

Salmonella typhimurium LT2

Salmonella typhimurium LT2

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.

1 (bases 1 to 22411)

McClieand,M., Sanderson,K.E., Spieth,J., Clifton,S.W., Latteille,P., Courtney,L., Potwolk,S., Ali,J., Dante,M., Du,F., Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A., Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Flores,L., Miller,W., Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.

Complete genome sequence of Salmonella enterica serovar Typhimurium LT2

Nature 413 (6858), 852-856 (2001)

21534948

11677609

2 (bases 1 to 22411)

The Salmonella typhimurium Genome Sequencing Project

Direct Submission

Submitted (29-MAR-2001) Genome Sequencing Center, Department of

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; <http://www.tigr.org/sofflab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
Reguondb;
http://kinich.cifn.unam.mx:8850/db/reguondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES

SOURCE

Location/Qualifiers

1..22411

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

434..1308

/gene="phoH"

/note="synonym: STM1126"

434..439

/gene="phoH"

/note="putative RBS for phoH; Reguondb:STMS1H001398"

454..1308

/gene="phoH"

/note="similar to E. coli PhoB-dependent, ATP-binding pho
regulon component; may be helicase; induced by P
starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
aa), 92% identity in aa 71 - 354"

/codon_start=1

/transl_table=1

/product="PhoB-dependent ATP-binding pho regulon
component"

/protein_id="AAL20057.1"

/db_xref="GI:16419642"

/translation="MCANRVQENAGGVYQLKLSNGRKAVIKAREAKRVLRRDS
RSHKREESVTSLVQMGVGAELIMARDSDPTSPVKAENEQAHYLAINDSKQLIFAT
YDVLKRLGAFMOCYLRPEIKVEIAPPAVGRGTENAVVILDEAONVTAQMKRP
LTRGENVTYVINDITQCCLPRGVRSGSLALRFEDEMGVIVHKNKDCVTSALC
QRTIHAYS"

complement(1414..2303)

/gene="STM1127"

complement(1414..2295)

/gene="STM1127"

/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%

identity in aa 15 - 264"

/codon_start=1

/transl_table=1

/product="putative transcripional regulator"

/protein_id="AAL20058.1"

/db_xref="GI:16419643"

/translation="MEQPPRLKPKGLDITGAMQKULTRASQRIAOYIAPPOVTO
SSIADISRETQAGEATVIRFRTLGKGFQDFKNDLIELATTSDDSSPLDVEE
SDDHAIGLKQNTISNVLETNLDMQVLGVVDALRHCHSVYIFGVSGGTTALD
MKHLMRIIGLRGDAVSNHFMVQATLLKADVAMGVSHSGTSPETVHSIRARQGA

RBS

TTVATITNLGSPICEADPCLINGNRQMLQGDIGTKAQLFVFDDLTLLVOSSPE
QARESKRTMNAIDMTX"
complement(2298..2303)

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

complement(2298..2303)
/gene="STM1127"
/note="putative RBS for STM1127; Reguondb:STMS1H001399"
complement(2580..4088)
/gene="STM1128"
complement(2580..4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=1
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MTHSRGIVYVLYVFGYLAAMLVGVFSRROKADDPYGGGR
VPGMAGVSPATTLISITFMSIPAKAFSDMPFIIGOVLAIALPLVFPYVPPRK
LKTSAEYELARDVRCRLPASMPHLPHIGRAIITPLVLAIRPAIDPILVIL
LISVCIITVMGIGBGVITVDVYQGLLSGSAITLITVCLKQGGIDELFTVQQA
DKFPPATQFPMSTESTVPMVLMIGFLPANTQOFTASODVQVRYVTDSIEBTKTLLT
NAXLAVAVIPVFPAIGSALPVYQOHPQLPAGFNTGILPLFVTEMVPGIACILIA
ALFAAQSSISSINSISCSFNSDIYORLSHKRTPEMRKIAKLVLVAGLISAS
VVLVMADESEIWDAPNSLIGLMGPMGLMGLGIFPRANAGSAVLGIITISVITVLA
RYATDLPFPYGVGSLSVYISGVTRAPLRAPPLTDEKPEKTL"
complement(4083..4088)
/gene="STM1128"
/note="putative RBS for STM1128; Reguondb:STMS1H001400"
complement(4413..5099)
/gene="STM1129"
complement(4413..5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSILARLEQSVHENGGLIVSCQVPSPMDKPEIYVAAQAAS
AGAVAVRIGENRTVRPHLSVPIIGIRDLGSRVTRPYQVDDALAAQAGADI
APASPRSRVDDLSLTRRLHGLIAMAQCSVNEGISCHOKIEITGTLISGVTP
ITPVEDLAVTQLSHAGRCVIAEGRYNTPALANALIEHGMATVTSATIRIHICQ
WFSHAVER"
complement(5094..5099)
/gene="STM1129"
/note="putative RBS for STM1129; Reguondb:STMS1H001401"
5599..6759
/gene="STM1130"
5599..6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQKRPKMMQMTALCSVAITTSQVATLLPDIPBLKNGG
AIDNGVITVGLGTAGTSWYKIDKQKHQWETIKPFGCARQESVFLDELIVYVG
GVKQKSESPLOVYSVYKTSVPKNTQKQDTSVGLTGTGKLETNVLITGVN
EHI.FDKYFIDIAAADADESKNVIYVFKPAKADYEFNIVFLYNAKENTKNAAGL
PGAGTSSSVMMNNFLMLINGELKFLRDVLYRAMMDNDKLTWLNKSQLPSPGSG
QOAGLGAFFSGYSHGVLLVGGANFPKALQNTNGKRYSHGINKKRDVYGLINGH
QOAGLGAFFSGYSHGVLLVGGANFPKALQNTNGKRYSHGINKKRDVYGLINGH
6793..7497
/gene="STM1131"
6793..6798
/gene="STM1131"
/note="putative RBS for STM1131; Reguondb:STMS1H001402"

CDS

6805..7497
 /gene="STM1131"
 /note="similar to E. coliorf, hypothetical protein
 (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 268
 identity in aa 4..239"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAL20062.1"
 /db_xref="GI:16419647"
 /translation="MKINKYLGMVFSFYQAATLDYRHYADRTIRINDRAI
 IEKLPNGIIGFYVDASVKSQGVGDGKHSLDVLNANATEIGVSNYKTDIVLQPGFI
 EESGPDTSIYKRYLGYNGYNDGSGVGMGRVYDARKTANYSDEKTNRRDTYIGVF
 DELKLENYFTWMDSDQIKFKDKNTNENHNAIAWKLNKSFYVEGVNAVVRNNTDER
 QTRVRLQVHP"
 7768..9060
 /gene="STM1132"
 7768..7773
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
 7780..9060
 /gene="STM1132"
 /note="similar to E. coli putative transport protein
 (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 578
 identity in aa 25..418"
 /codon_start=1
 /transl_table=11
 /product="putative sugar transport protein"

Query Match 98.2%; Score 448; DB 1; Length 22411;
 Best Local Similarity 98.9%; Pred. No. 4,3e-118;
 Matches 451; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

1 ATGAACCTTTTAAAGTGGAGCATTCGAGCAATGCTGTTCTGGACAGTCTGGCT 60
 Db 17769 ATGAACCTTTTAAAGTGGAGCATTCGAGCAATGCTGTTCTGGACAGTCTGGCT 17828
 Qy 61 GGGCTGCTTCCCAATGGGGCGCGCGGTATCTAATACGGCGCGCGCAATAGTTCGGC 120
 Db 17829 GGGCTGCTTCCCAATGGGGCGCGCGGTATCTAATACGGCGCGCGCAATAGTTCGGC 17888
 Qy 121 CCGGACTCAAGCTTGAGCATTTATGATGATGCTTCCGCTAAGCGCGCTTGGCTCGCA 180
 Db 17889 CCGGACTCAAGCTTGAGCATTTATGATGATGCTTCCGCTAAGCGCGCTTGGCTCGCA 17948
 Qy 181 ACCGATGCCCTTAATCTGAACGACCACTTACCGAGCGGTATGTTAAGCGCGCGAT 240
 Db 17949 ACCGATGCCCTTAATCTGAACGACCACTTACCGAGCGGTATGTTAAGCGCGCGAT 18008
 Qy 241 GTAGGCGAGGCTGGCATTAATAGTATGATGATGATGATGATGATGATGATGATGAT 300
 Db 18009 GTAGGCGAGGCTGGCATTAATAGTATGATGATGATGATGATGATGATGATGAT 18068
 Qy 301 GCCACATCGACGACGTGAAGCTTAAATCTCGATATTAATCTCGGCGCATACGGCGGT 360
 Db 18069 GCCACATCGACGACGTGAAGCTTAAATCTCGATATTAATCTCGGCGCATACGGCGGT 18128
 Qy 361 AATAACGCCGGGCTGTTAATCAGACCGCATCTGATTCAGACGCTATGATGATGATGAT 420
 Db 18129 AATAACGCCGGGCTGTTAATCAGACCGCATCTGATTCAGACGCTATGATGATGAT 18188
 Qy 421 GGTGGGCAACAGCGCAGCGCTAACAGATTA 456
 Db 18189 GGTGGGCAACAGCGCAGCGCTAACAGATTA 18224

RESULT 5
 AL627269 254050 bp DNA linear BCT 04-JUL-2003
 LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
 DEFINITION complete chromosome; segment 5/20.
 ACCSSION AL627269 AL513382
 VERSION AL627269.1 GI:16502231

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

TITLE

COMMENT

FEATURES

SOURCE

1. 254050
 /organism="Salmonella enterica subsp. enterica serovar
 Typhi"
 /mol_type="genomic DNA"
 /strain="CT18"
 /db_xref="taxon:90370"
 181..2793
 /gene="STY1078"
 /note="synonym: pepN"
 181..2793
 /gene="STY1078"
 /note="orthologue of E. coli pepN (AMPN ECOLI); Fastr hit
 to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
 /codon_start=1
 /product="aminopeptidase N"
 /protein_id="CAD08183.1"
 /db_xref="GI:16502232"
 /db_xref="GOA:O82770"
 /db_xref="SPRMBL:O82770"
 /translation="MTQOPQARYRDYAPDQITDIEFDLDAEKTVAIAOAVR
 HGAADPRLDDELTLVSIHVNDAPMTAYKEEBAALISLPFRITLRYNEISPA
 NTALELYSGDALCTOCBAEGRHITVYLRPDVLAFTKTLADSKYFPLISNGA
 RVAQGLENGKRWVOMODPEKPYCLFALVAGDPVLAFTTTSGBVALLELVDS
 LARTDADKQVLDLERYTGHRYRNATGKNTVCTWDEFLSKGLTVFRDSESLD
 TILGRENPKQMLFERHDSATPCDFVOMADASNVDSLHFRWYSGSGPTIVY
 KDVINPEBQVTLTISQRTPATDAQEKQPHIFPALELYDNEBNVPLKGGHPVNA
 VNAVQAEQTEFEDNVYFQVPALCEFSAPVLSLEKYSQGLTFRLRHNRDPSRD
 MAOSLIATYIKLVNVAHQOQPLSLPVAVDAFPAYVLLDKTIDPALAAILTPSAE
 IAELEFVIDPLIAQVREALTRTIAELADEFLAIYVANNLDERVDHIGIKRTLN
 AGLRLTFGEETELANTLVSROYRANMTDIALALSAVAHQPCRDTLQGEYDRLN
 ODGLVMKFTLOSSTPAENVLETVRSLLKRSFSMSNPRIISLIGAFSGNPAAFH
 AODSGXQVLEVMETDLANSRNPVARSRLISPLRLKRYDDKOEKRAALEQKLEN
 LSGDIYEKITKALA"
 532..1332
 /gene="STY1078"
 /note="Pfam match to entry PF01433 peptidase_M1, peptidase
 family M1, score 245.10, E-value 1.9e-84"
 1060..1069
 /gene="STY1078"

misc_feature
 misc_feature

gene /note="P800142 Neutral zinc metalloproteases, zinc-binding region signature"
3001. .4011
/gene="STY1079"
/note="synonym: PYRD"
3001. .4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q82759"
/db_xref="SWISS-PROT:Q82759"
/translation="MYPPVPRKALFQDLRERAHETPOQLPRITGPEALVPOKVT
KVYTCMGLFKRPLGLAAGLDKDGSCDALAMGFGSLKETVTPRPQGRKRLFR
LVDAAGLIRAFNFDVNLVFNKKAHPGILGINGRKDTPEVNGKADYLTICME
KYVAAGYAIRNIISSPNTPEGLTLOYGADLLDLTLAIKIKONDLQVTHKKVPAVNI
ADLCEEBELIIVQADSLRHNIDVGIATVTTDLDRSLVQGMKCOQTGSLGSPDLKST
EIRRLSELKQGLPIIGVGIDSVIAREKILAGATLVQYVSFIRKGPLLIEITYT
H"
misc_feature 3136. .3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHodehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
misc_feature 3238. .3297
/gene="STY1079"
/note="P800911 Dihydroorotate dehydrogenase signature 1"
3874. .3936
/gene="STY1079"
/note="P800912 Dihydroorotate dehydrogenase signature 2"
4174. .4719
/gene="STY1080"
4174. .4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q82758"
/db_xref="EMBL:Q82758"
/translation="WMRIKPDNMKMYVEEDHDMMLDLANGMLFRSFRSKRLTPDA
FCPTAFCDVDAALYFSPFEKCRDELTEQGAELVINALVAIRYLRKQMSKMHFVNH
GEMTPPGVDAASVMSDTAEQVNLVVEPENNALCILAPGVVINGRTWQJGDAIK
IMNDLRKQVCHSPLEQAV"
complement(4716. .5825)
/gene="STY1081"
complement(4716. .5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82757"
/db_xref="SPTREMBL:Q82757"
/translation="WATLSRLFIHVSMRGICGLTHALDISGLAFDRILFMVTSQDT
FITAQFPQMAWRTPSPLDHGLJHLIAPDGSALVRFPDPPQDAFTVWNGHTFARVA
PTAIQMLSGFSSRDVQCLVQGPOLTRVKKHNAVPLGADGYLLTTFASRLDQO
RCPAQVOMGSRPNLVNIGVAVAMEBDSKVLRIQDVI EDVVKPCSRCLPTVSEKGO
KHPSEPLATLQAPRTADNGDVRCGNLTARNGLRVDEVGLIATAPAKAYGATT
LDISTPEPKHDSYTTIDMOCOTRCANNOVLRLQENQGRIRIYSCRAIGCGCRRIR
LDEGVSPDKKSAWMDSDTILISCSQVPRKTLRLN"
complement(4737. .4955)
/gene="STY1081"
misc_feature

Query Match	Best Local Similarity	Matches	450; Conservative
97.9%; Score 446.4; DB 1; Length 254050;	98.7%; Pred. No. 1,5e-117;	0; Mismatches 6;	Indels 0; Gaps 0;
1	ATGAATCTTTAAAGTGCAGCATTCGCGACATCGTATGTTCTGCGAGTGTCTGAGCT	60	
88906	ATGAATCTTTAAAGTGCAGCATTCGCGACATCGTATGTTCTGCGAGTGTCTGAGCT	60	
61	GCGCTCGTTCCACATATGGGCGGCGCGGTATCATATCGCGCGGCAATATGTTCCGCG	120	
88966	GCGCTCGTTCCACATATGGGCGGCGCGGTATCATATCGCGCGGCAATATGTTCCGCG	120	
121	CCGAGCTCAACGTTGACATTTATTCAGTACGTTCCGCTTAAAGCGCTTCTCTGCGAA	180	
89026	CCGAGCTCAACGTTGACATTTATTCAGTACGTTCCGCTTAAAGCGCTTCTCTGCGAA	180	
181	AGCGATCCCGTAAATCTGAAACGACATTAACCGACAGCGGTTATGTTAAGCGCGCGAT	240	
89086	AGCGATCCCGTAAATCTGAAACGACATTAACCGACAGCGGTTATGTTAAGCGCGCGAT	240	
241	GTAGGCCAGGGTCCGATTAATAGTACTATGTTAACTGACTCAGATGTTTCAGAAATAAT	300	
89146	GTAGGCCAGGGTCCGATTAATAGTACTATGTTAACTGACTCAGATGTTTCAGAAATAAT	300	

```
Oy 301 GCCACCATCGACGATGAGAAAGCTAAATCTCCGATATTACTGTGGCCATATACCGCGGT 360
Db 89206 GCCACCATCGACGATGAGAAAGCTAAATCTCCGATATTACTGTGGCCATATACCGCGGT 89265
Oy 361 AATAACCGCCGCGGTATATCAGACCGCATCTGATCCAGCGTAATGGTGCCTCAGGTT 420
Db 89266 AATAACCGCCGCGGTATATCAGACCGCATCTGATCCAGCGTAATGGTGCCTCAGGTT 89325
Oy 421 GCTTTTGGCAACGACCGCGCTAACGATTTAA 456
Db 89326 GCTTTTGGCAACGACCGCGCTAACGATTTAA 89361

RESULT 6
LOCUS AE016840/c 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
ACCESSION AE016840 AE016613
VERSION AE016840.1 GI:29137797
KEYWORDS
SOURCE
ORGANISM Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 301983)
Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.,
Comparative Genomics of Salmonella enterica Serovar Typhi Strains
CT18 and CT18
JOURNAL Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE 12644504
PUBMED 22531367
REFERENCE 2 (bases 1 to 301983)
AUTHORS Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.,
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 301983
/organism="Salmonella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
/notes="corresponds to STY1225 from Accession AL513382:
Salmonella typhi CT18"
CDS
/protein_id="AA069359.1"
/db_xref="GI:29137798"
/translation="MSDEKSLATRNCKYHIVAPRYRQAFYGEKRAVGSILRLK
CEKAVYRILAECCADHIMLEIPKMSVSSFMGLGKSGIMTYRGDLAKRYKN
REWRCGYVYDVGKNTAKIDYIKHQLEEDMGESLIPYPSFTGRK"
complement(844..1797)
/gene="fliG"
/locus_tag="t1736"
complement(844..1797)
/locus_tag="t1736"
/notes="corresponds to STY1223 from Accession AL513382:
Salmonella typhi CT18"
CDS
/protein_id="AA069360.1"
/db_xref="GI:29137799"
/translation="WRISTOMYEQNNSGITNSQAEWMKLEGMSGKEKVTNPDPI
AASQAVYFQAOAONSOYALARTFATDQKSLSESVLSQVTTAIOAEKIVVAGNGTL
TQDDRAILADLQIRIQALNLSNDNGRYIFAGYKTEAAPPDQATGCGYGEKSV
DVEKKAAPAVIGHTGAQIPNSITNSAVPEPDSDESKULFWMLDPAIALATPVEGN
NLVDVWNVSIVSSVYMOQALQASVKTFTDMQMSLFGQNR"
complement(1812..3473)
/gene="fliG"
/locus_tag="t1737"
complement(1812..3473)
/gene="fliG"
/locus_tag="t1737"
complement(1812..3473)
/locus_tag="t1737"
/notes="corresponds to STY1222 from Accession AL513382:
Salmonella typhi CT18"
CDS
/protein_id="AA069361.1"
/db_xref="GI:29137800"
/translation="MSSLINHAMSGLANAQAALNTVSNINNNVAVGTRQTTIAQA
NSTIAGGKWTIGVYVSGVOREYDAFTINQLRGAQONOSGLTTRBQMSITDNLADK
SSLSGSLQSFPTSLQTLVSNADPEARQALIGRAGLVNPFRTDQYLDQDKQNI
AIGSSVAQINNVAKQIANLNDQISRMGVGASPNLLDQORDVSELNKIVGEVS
VODGGTYNLVTWANGYTLVQSGTARQLAAVPSADPTTTAVYDEAAGNIEIEKLN
TSGISGLTIFPSODLDQTRNTLQALAPADRAHQTKYADGNGKQDFSIQSPV
VYSNNMADKTVSLTAKYVDSYTKQAQDYKI VFDGTDQVTRADMTTFATQDAQK
LEIDELKTYGTCGQXNDLFLKPSNAIVDMVKYINAEIIMASESKLDPVDVGD
SDNRNGQALILQNSNVGKTFNDAYATLVDSGNTSTLKTSSSTQANVVKQLYK
QQQSYSGVNLDEEYGNLQRYQOYYLANAQLQANALFDALNLR"
complement(3538..4488)
/gene="fliG"
/locus_tag="t1738"
complement(3538..4488)
/gene="fliG"
/locus_tag="t1738"
complement(3538..4488)
/locus_tag="t1738"
/notes="corresponds to STY1221 from Accession AL513382:
Salmonella typhi CT18"
CDS
/protein_id="AA069362.1"
/db_xref="GI:29137801"
/translation="MTGDKRLASAAQAQSLNELKAKAGDPAANIRPVAROVGME
TMPADAPVPLKFSLETVNSYQNALQOLYKAIPTKPPSSDAPLPGSDSKDFARLS
LPRKLSQSGVPHHLILAQALLESQMGORILRENGRPSYNNVGUYATASMKPQVE
ITTYEYENGEAKKAVAKFRVSSYLEALSDVALLTTRPRAVTTAATAEQGAVALQ
NAGVATDPNYPARKLASMIQOLKAMESEKSKYSANLNLTF"
complement(4488..5591)
/gene="fliG"
/locus_tag="t1739"
complement(4488..5591)
/gene="fliG"
/locus_tag="t1739"
complement(4488..5591)
/locus_tag="t1739"
/notes="corresponds to STY1220 from Accession AL513382:
Salmonella typhi CT18"
CDS
/protein_id="AA069363.1"
/db_xref="GI:29137802"
/translation="MYVFKALAGIVLAVVTLAHERIDLTSGVGRNSLITGGLV
VGLDGTGDTQTFQTFOTLNNMLSGLGTVPGTMMQKRAAVVNTASTYPPFARQG
QTIQVIVVSMGNAKSLRGGLTMTPLKGDQSVYALAQSNILVGAAGASAGSSVQVN
QLNAGRTNGAIIERLPTQFGAGNTINQLNDEDFMAQITTDAINRARGSATL
DARTVOYRVASGNSQVRPLADIQNMEVAVTQDAKVVINSRGSVNAEVEVTLDSCA
VAGSNLSVTNRQLNVAQNPITPFGGQIVVTQTOIDLRQSGSLQSVSSANLSNV
```

gene
BALNAGATPMDLMSILQSMOSAGCLRAKLEII"
complement(5597. .6295)
/gene="f1gH"
/locus_tag="c1740"
complement(5597. .6295)
/gene="f1gH"
/locus_tag="c1740"
/note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MQKYLHAYPVMALMTATLTCGAMIPAKPLVQGAATTQPIPGPV
EVANGSIPOSQPINYGYPLFEDRPNIDITLTIIVQENVASAKSSNAASRDKT
SFGPDYPRVYIOGLFGNSRADMEASGSGFNGKGANASNTFSGLTITVQVLANGN
LHVVEKQIAINOGTEFIRFSGVVPRTISGNSVPTQVADARIEYVNGYINBAON
MGWLORFPLNLSPM"
complement(6350. .7132)
/gene="f1gC"
/locus_tag="c1741"
complement(6350. .7132)
/gene="f1gC"
/locus_tag="c1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein f1gC"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISLWTAKTGLDAQQTNDVIANNLANVGTNGCKRQRAVFEDL
LYQIRQGAQSSQETLPSGLQIGTVRPATERLHSGNLSTQNNKDAIKQGF
FQWMLPDGSAVTRDGSFQVNDQGLTVAGSFQVOPAITIPANALSTIGRDGVSVT
QOQGAAPVOGQLMTTFMNDTGLSEIGENIYIETOSGAPNESFPGINGAGLKYGY
VETSNNVAEEVNMIVORAYEINSKAVSTTDQWLKRLTQL"
complement(7146. .7901)
/gene="f1gF"
/locus_tag="c1742"
complement(7146. .7901)
/gene="f1gF"
/locus_tag="c1742"
/note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein f1gF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHALITRMGAASQTLANQAVTASNLANASTPGFRQAOLALRAV
PVDGLSLATRLTLVASTPGADMTQGLDYTSRPLDVALQDQGLVQADAEQYTRN
GNIQVPGTQGLTIQHPVIGSGPITVEGSEITADGITSALNPGSPPTVAEVR
LKLVAEGNEVORSDDLGFLRLAEQAERGVLADPSIRIMSGVLBSNVKPYEAMT
DMINARREFOMKVTISVDNEGRANLLMS"
complement(7922. .9133)
/gene="f1gE"
/locus_tag="c1743"
complement(7922. .9133)
/gene="f1gE"
/locus_tag="c1743"
/note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein f1gE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFSQAVSGLMAATNDVIGNNANSATGFKSGTASPADMPA
GSKVGLGVKAVGITDDPFDGTTTNGRLDVAISONGFFRLVDSNGSVFYSRNGQFKL

gene
CDS
Query Match
Best Local Similarity

97.9%; Score 446.4; DB 1; Length 301983;
98.7%; Pred. NO. 1.5e-117;

Matches 450; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGAGTCTTGCT 60
DB 37310 ATGAAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGAGTCTTGCT 37251
QY 61 GGGCGTCGTTCCAAATGGGGGGGGGGGGGTATATATACGGCGGGGGAATGTTCCGGC 120
DB 37250 GGGCGTCGTTCCAAATGGGGGGGGGGGGGTATATATACGGCGGGGGAATGTTCCGGC 37191
QY 121 CCGGACTCAAGCTTGAAGATTATACAGTCCGTTCCGTTACGCTCCGTTCTTGCAA 180
DB 37190 CCGGACTCAAGCTTGAAGATTATACAGTCCGTTCCGTTACGCTCCGTTCTTGCAA 37131
QY 181 AGCGATGCCGTTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAACGGCCCGAT 240
DB 37130 AGCGATGCCGTTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAACGGCCCGAT 37071
QY 241 GTAGGCCAGGGTGGGGAATATAGTACTATTTGAACGACTCAGAAATGTTTCAGAAATAT 300
DB 37070 GTAGGCCAGGGTGGGGAATATAGTACTATTTGAACGACTCAGAAATGTTTCAGAAATAT 37011
QY 301 GCCACCATCGACAGTGGAAAGCTAAAACTCCGATATTACTGTCCGCCAATACGCGCGT 360
DB 37010 GCCACCATCGACAGTGGAAAGCTAAAACTCCGATATTACTGTCCGCCAATACGCGCGT 36951
QY 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCACGCTAATGTCGTCAGGTT 420
DB 36950 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCACGCTAATGTCGTCAGGTT 36891
QY 421 GGTTCGCAACACGCGCAGCGCTTACAGATTAA 456
DB 36890 GGTTCGCAACACGCGCAGCGCTTACAGATTAA 36855

RESULT 7
STAGFBA
LOCUS
DEFINITION STAGFBA 1048 bp DNA linear BCT 26-JAN-1998
ACCESSION AJ000514
VERSION AJ000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
TITLE Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE
AUTHORS Sukupolvi,S.S.
TITLE Direct Submersion
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinanmyllykatu, 20520, FINLAND
FEATURES
source
1. 1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14. .553
/gene="agfB"
14. .553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="AgfB protein"
/protein_id="CAA04150.1"

```

/db_xref="GI:2275120"
/translation="MKRKLFLMMLTILGPIATATNYDLAREYNFAVNEISKSPNO
AATIGVDIENSARVHESKLSVISCERGNMNAKIDQANVNAFATIEGTGANDA
SISQSAVGNSSAISAIIQKSGNKANITQYGTQKTAIVVOKOSHMAIQANTQYGTQ
KTAIVVOKOSHMAIRVYGR"
593..1048
/gene="agfa"
593..>1048
/gene="agfa"
/codon_start=1
/translation="Agfa protease"
/protein_id="CAA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:O33802"
/translation="MKLTKVAFPAIVVSGAVAVPQWGGGNGHNGSSGPDST
LSIYQYGSNMAIALQSDARKSEFTITQSGYNGADVGGADNSTELIENGPRNNAT
IDQMAKNSDITVQYGGNNALVNTASDSVVRQVGFNNAPANQYN"

```

ORIGIN

```

Query Match      96.6%; Score 440.6; DB 1; Length 1048;
Best Local Similarity 98.0%; Pred. No. 4.9e-116;
Matches 446; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 1 ATGAACCTTTAAAGTGGAGCATTTCGACGCAATCGTACTTCTGGAGTGTCTGGCT 60
Db 593 ATGAACCTTTAAAGTGGAGCATTTCGACGCAATCGTACTTCTGGAGTGTCTGGCT 60
QY 61 GGGCTGCTTCACAAATGGGGCGGGCGGCTAATCAATACGGGGCGGCAATGTTCCGGC 120
Db 653 GGGCTGCTTCACAAATGGGGCGGGCGGCTAATCAATACGGGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAAGTGTGACATTTATCACTAGACGCTTCGCTAAGCGCTGCTGCTGCA 180
Db 713 CCGGACTTCAGTGTGACATTTATCACTAGACGCTTCGCTAAGCGCTGCTGCTGCA 180
QY 181 AGCGATGCCGTAATCTGAACGACCATTAACCGAGCGGCTTATGTAACGGCGCGAT 240
Db 773 AGCGATGCCGCTAAATCTGAACGACCATTAACCGAGCGGCTTATGTAACGGCGCGAT 240
QY 241 GTAGGCGCAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 833 GTAGGCGCAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACAGTGAACGCTAAAACTCCGATATTAATGTCGGCGCAATACGGCGGT 360
Db 893 GCCACCATGACAGTGAACGCTAAAACTCCGATATTAATGTCGGCGCAATACGGCGGT 360
QY 361 AATAAGCGCGGCTGTAATCAAGCGCATCTGATTCAGCGTAATGTCGGCGCAAT 420
Db 953 AATAAGCGCGGCTGTAATCAAGCGCATCTGATTCAGCGTAATGTCGGCGCAAT 420
QY 421 GGTTCGCAACAAAGCGCAGGCTAATCAAGATTA 455
Db 1013 GGTTCGCAACAAAGCGCAGGCTAATCAAGATTA 1047

```

```

RESULT 8
LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 56 from patent US 5635617.
ACCESSION 144908
VERSION 144908.1 GI:2469621
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 361)
TITLE Doran,J.L., Kay,W.W., Collinson,S. Karen. and Clouthier,S.C.
METHODS Methods and compositions comprising the agfa gene for detection of
JOURNAL Patent: US 5635617-A 56 03-JUN-1997;

```

```

FEATURES
source Location/Qualifiers
1..361
/organism="unknown"
/mol_type="unsigned DNA"

```

ORIGIN

```

Query Match      72.8%; Score 331.8; DB 6; Length 361;
Best Local Similarity 99.4%; Pred. No. 1.2e-84;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY 64 GTGCTTCACAAATGGGGCGGGCGGCTAATCAATCAATCAATCAATCAATCAATCAATCAAT 123

Db 1 GTGCTTCACAAATGGGGCGGGCGGCTAATCAATCAATCAATCAATCAATCAATCAATCAAT 123

QY 124 GACTCAACGTTGAGCATTTATCACTAGTACGTTCCGCTAAGCGTGGCTTCTCTGCAAAAC 183

Db 61 GACTCAACGTTGAGCATTTATCACTAGTACGTTCCGCTAAGCGTGGCTTCTCTGCAAAAC 180

QY 184 GATGCCCGTAATCTGAACGACCATTAACCGAGCGGTTATGTAACGGCGCGCATGTA 243

Db 121 GATGCCCGTAATCTGAACGACCATTAACCGAGCGGTTATGTAACGGCGCGCATGTA 180

QY 244 GGCAGAGGTGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 303

Db 181 GGCAGAGGTGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240

QY 304 ACCATGACACAGTGAACGCTAAACCTCCGATTAATGTAATGTAATGTAATGTAATGTAAT 363

Db 241 ACCATGACACAGTGAACGCTAAACCTCCGATTAATGTAATGTAATGTAATGTAATGTAAT 300

QY 364 AACGCCGCGCTGTTAATCAAGACCGCATCTGATTC 398

Db 301 AACGCCGCGCTGTTAATCAAGACCGCATCTGATTC 335

```

RESULT 9
LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter sp. Fec2 csb gene, csbA gene and csbD gene.
ACCESSION AJ515700
VERSION AJ515700.1 GI:11790491
KEYWORDS csbA gene; csbB gene; csbD gene; curlin-csbA protein; nucleation
component of curlin monomers; regulatory protein.
SOURCE
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Citrobacter.
REFERENCE
1 Zogaj,X., Bokranz,W., Nimmz,M. and Romling,U.
AUTHORS Production of Cellulose and Curli Fimbriae by Members of the Family
JOURNAL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
REFERENCE Infect. Immun. 72 (7), 4151-4158 (2003)
AUTHORS 2 (bases 1 to 2889)
TITLE Romling,U.
JOURNAL Direct Submision
SUBMITTED (11-NOV-2002) Romling U., Microbiology and Tumorigenology
JOURNAL Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
source Location/Qualifiers
1..2889
/organism="Citrobacter sp. Fec2"
/mol_type="genomic DNA"
/isolate="Fec2"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
/complement(212..862)
/gene="csbD"
/complement(212..862)
/gene="csbD"
/function="necessary for transcription of the csbA
operon"
/codon_start=1

```

gene

CDS

```

/codon_start=1

```

```

/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MFNEFHSIHGHTLLITKPSLOATALLQHLKOSIALSGKLANIO
RSLDDISSCIVLDMEADKLIHYQDNLRSKNNIKTLILTPDIPYDIENMP
HINGVFAVDEEERVNGLOGLRGECEFSOKLASYLTHSGNRYNSSEFALLTHE
KEILNKLRIGASNEIARSLFISENTVKTHLYNLFKIAVNRTOAVSANDLNR"
1623..2078
/gene="csaB"
/locus_tag="2078"
/cds_id="2078"
/function="nucleator"
/codon_start=11
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKKKLFFMFLTVIGAPGASATSYDLAHSFNPVNELSKSFN
OALIGOVGTNSAKMRQEGSKLSVISOEGSNRAVDSGAINPFIYIAGSHSND
SISQSNYNTAMITIQSGNKNANTIOYGTOKTAVVOROSOMAIRVOR"
2119..2571
/gene="csaA"
/locus_tag="2571"
/cds_id="2571"
/function="curli subunit"
/codon_start=11
/transl_table=11
/product="curlin-csaA protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLIQVAFPAIYVSGSALAGVPOMGGGGGGSSSGSESTL
SIYSGNNALAIQSDARKSDTIHQGRNGADVGGSDNSTIDITQNGFKNNTI
DOMGKNSDITVSOYGGHNAALVNOTASDSVLVHVGFGNNAATANQY"
ORIGIN
Query Match      68.1%; Score 310.4; DB 1; Length 2889;
Best Local Similarity 81.6%; Pred. No. 2.1e-78;
Matches 372; Conservative 0; Mismatches 81; Indels 3; Gaps 1;
QY 1 ATGAACCTTTAAAGTGGCAGATTGCGACAAATCGTAGTTTCTGGCAGTCTTGCGCT 60
DB 2119 ATGAACCTTTAAAGTGGCAGATTGCGACAAATCGTAGTTTCTGGCAGTCTTGCGCT 2178
QY 61 GGGCTGTTCCAAATAGGGGGGGGGGCTATATAACGGCGCGCAATATGTTCCGGC 120
DB 2179 GGTTCGTTCCGCAATAGGGGGGGGGGCTATATAACGGCGCGCAATATGTTCCGGC 2235
QY 121 CCGGACTCAAGTGGAGATTATCAGTACGGTTCGCTAAAGCTGGCGCTGCTCTGCA 180
DB 2236 CCGGACTCAAGTGGAGATTATCAGTACGGTTCGCTAAAGCTGGCGCTGCTCTGCA 2295
QY 181 AGCGATGCCGTAAATCTGAACGACCAATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
DB 2296 AGCGATGCCGTAAATCTGAACGACCAATTAACGAGCGGTTATGTTAAGCGCGCGAT 2355
QY 241 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTCACTCAAGTGGTTTCAAAATAT 300
DB 2356 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTCACTCAAGTGGTTTCAAAATAT 2415
QY 301 GCCACATCGACGAGTGAACGCTAAACTCCGATATTAATGTCGGGCAATACGGGGGT 360
DB 2416 GCCACATCGACGAGTGAACGCTAAACTCCGATATTAATGTCGGGCAATACGGGGGT 2475
QY 361 AATAAGCCCGGCTGTTAATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGGTT 420
DB 2476 CATTAACCGCGCATCTGTTAATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGGTT 2535
QY 421 GGTTTGGCAACAACGGCAGGCTAACCAATATTA 456
DB 2536 GGTTTGGCAACAACGGCAGGCTAACCAATATTA 2571

```

RESULT 10

CFR515701

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

CFR515701

Cytobacter freundii csaB gene, csaA gene and csaD gene.

A515701

A515701.1

GI:31790495

csaA gene; csaB gene; csaD gene; curlin-csaA protein; nucleation

component of curlin monomers; regulatory protein.

Cytobacter freundii

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Cytobacter.

1

Zogaj, X., Bokranz, W., Nimtz, M. and Romling, U.

Production of Cellulose and Curli Fimbriae by Members of the Family

Enterobacteriaceae Isolated from the Human Gastrointestinal Tract

Infect. Immun. 72 (7), 4151-4158 (2003)

2 (bases 1 to 2920)

Romling, U.

Direct Submission

Submitted (12-NOV-2002)

Romling U., Microbiology and Tumoriology

Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN

Location/Qualifiers

1..2920

/organism="Cytobacter freundii"

/mol_type="genomic DNA"

/isolate="Fec4"

/isolation_source="faeces"

/specific_host="Homo sapiens"

/db_xref="taxon:546"

/country="Germany"

complement(211..861)

/gene="csaD"

complement(211..861)

/gene="csaD"

/function="necessary for transcription of the csaA

operon"

/codon_start=1

/transl_table=11

/product="putative regulatory protein"

/protein_id="CAD56673.1"

/db_xref="GI:31790496"

/translation="MFNEFHSIHGHTLLITKPSLOATALLQHLKOSIALSGKLANIO

RSLDDISSCIVLDMEADKLIHYQDNLRSKNNIKTLILTPDIPYDIENMP

HINGVFAVDEEERVNGLOGLRGECEFSOKLASYLTHSGNRYNSSEFALLTHE

KEILNKLRIGASNEIARSLFISENTVKTHLYNLFKIAVNRTOAVSANDLNR"

1624..2082

/gene="csaB"

1624..2082

/gene="csaB"

/function="nucleator"

/codon_start=1

/transl_table=11

/product="nucleation component of curlin monomers"

/protein_id="CAD56674.1"

/db_xref="GI:31790497"

/translation="MKKKLFFMFLTVIGAPGASATSYDLAHSFNPVNELSKSF

NOALIGOVGTNSAKMRQEGSKLSVISOEGSNRAVDSGAINPFIYIAGSHSND

SISQSNYNTAMITIQSGNKNANTIOYGTOKTAVVOROSOMAIRVOR"

2123..2572

/gene="csaA"

2123..2572

/gene="csaA"

/function="curli subunit"

/codon_start=1

/transl_table=11

/product="curlin-csaA protein"

/protein_id="CAD56675.1"

/db_xref="GI:31790498"

/translation="MKLIQVAFPAIYVSGSALAGVPOMGGGGGGSSSGSESTL

SIYSGNNALAIQSDARKSDTIHQGRNGADVGGSDNSTIDITQNGFKNNTI

DOMGKNSDITVSOYGGHNAALVNOTASDSVLVHVGFGNNAATANQY"

ORIGIN


```

misc_feature      /note="transcriptional start site for csgBDFG operon"
3140
gene              /note="transcriptional start site for csgBA operon"
3233..3688
CDS               /gene="csgB"
3233..3688
/gene="csgB"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAA62281.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translation="MKKLLPMLTTIGAPITAAAGYDLANSEYNPAVELSKSEN
QAALIGAGTNNASQLRQSGSKLAIVAAQESSNRAKIDQGDYNYLYIDAGSANDA
SISQAGYNTAMIIQKSGNKNANITQYQKTAIVOROSOMAIRVQR"
3729..4184
CDS               /gene="csgA"
3729..4184
/gene="csgA"
/note="major subunit"
/codon_start=1
/transl_table=11
/product="curlin/CsgA protein"
/protein_id="CAA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="MKLLKVAIAIAIVFSSGALAGVVOYGGGNGGGGNNSGPNS
LNIYQGGNSALALQTDARNSDLTIQHGNGNADYGGGSDSISDLTORGFNASAT
LDWNGKNSNTVQFGGNGAAVDQIASNSNVVTQVGRNNATAIQY"
3729..3788
81g_peptide      /gene="csgA"
4243..4575
/gene="orc"
4243..4575
/gene="orc"
/gene="orc"
/codon_start=1
/transl_table=11
/product="putative curli production protein"
/protein_id="CAA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="WNTLLALSLSSQITFTNTTQGGDVYTIIPVTLTQSLCVRQL
SLRSGSSGSGQTKQKTLSPANQPIALTKLSINISPDVRKIVVTVSDQSLSLQO
WPSSERS"
ORIGIN
Query Match      56.1%; Score 256; DB 1; Length 4680;
Best Local Similarity 72.8%; Pred. No. 1,1e-62;
Matches 331; Conservative 0; Mismatches 125; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGCAGCATTCGACCAATCGAGTTCGCGAGTCTGGCT 60
DB 3729 ATGAACTTTTAAAGTGCAGCATTCGACCAATCGAGTTCGCGAGTCTGGCA 3788
QY 61 GGCCTGTTCCACATGGGCGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
DB 3789 GGTGTTGTTCTCTGACGCGCGCGCGGTAAACACGCGTGTGCGGTAAATATAGCGGC 3848
QY 121 CCGGACTCAAGTGTAGATTTATCAGACGTTCCGCTACCGTGGCGTCTCTGCA 180
DB 3849 CCAATTTCTGAGTGAAACATTACAGTACGTTGCGGTAACTCTGCACTTCTGCA 3908
QY 181 AGCGATCCCCCTAATATGAAACAGCACTTACCGAGCGGTATGTGTAACGGCGCCGAT 240
DB 3909 ACTGATGCCCTAATCTGACTTATTAACAGATGCGCGCGGTAAATGTCGAGAT 3968
QY 241 GTAGCCAGGCTGGGATATGATCTATTTGAAGTGAAGTGTTCAGAAATAT 300
DB 3969 GTTGTGAGGGCTGATGACGCTCAATCATCTGACCCCAAGCTGCGTTCGTAAGACG 4028

```

```

QY 301 GCCACATCGACGATGGAACGCTAAAACTCGATATTTCTGCGGCATACGGCGCT 360
DB 4029 GCTACTCTGATGATGGAACGCTAAATTTCTAAATGACGTTAAACATCTGCGTCT 4088
QY 361 AATACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGGTAAAGTCCGACGTT 420
DB 4089 GGCACCGGTGCTGAGTGAACGACATCTGATCTCTCCGTCACAGTACCTCAGTT 4148
QY 421 GGTTTGGCAACAAACGCGCGCTTACCAAGTATTAA 456
DB 4149 GGTGTTGTAACAAACGCGCGCTCATCAGTACTAA 4184

RESULT 12
AE000205
LOCUS      Escherichia coli K12 MG1655 section 95 of 400 of the complete
DEFINITION genome.
ACCESSION AE000205.1 GI:1787265
VERSION    AE000205.1
KEYWORDS   Escherichia coli K12
SOURCE     Escherichia coli K12
ORGANISM   Escherichia coli K12
REFERENCE  Blatterer,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
AUTHORS    Riley,M., Collado-Vides,J., Glaesner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mu,B. and Shao,Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5311), 1453-1474 (1997)
9742617
MEDLINE    9278503
REFERENCE  2 (bases 1 to 10346)
AUTHORS    Blatterer,F.R.
TITLE      Direct Submission
JOURNAL    Submitted (16-JUN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            3 (bases 1 to 10346)
REFERENCE  Blatterer,F.R.
AUTHORS    Direct Submission
JOURNAL    Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            4 (bases 1 to 10346)
REFERENCE  Plunkett,G. III.
AUTHORS    Direct Submission
JOURNAL    Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using Genemark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications

```


and products: all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source

Location/Qualifiers

1. 10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"

/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
complement(<1..48)
/note="IS3"

<1..5

/note="factor Sigma70; predicted +1 start at 1094689"

5. 32 /note="factor Sigma70; predicted +1 start at 1094716"

69. 392 /gene="b1028"

69. 392 /gene="b1028"

/function="orf; Unknown"
/note="O107; This 107 aa ORF is 27 pct identical (0 gaps) to 58 residues of an approx. 1888 aa protein PAS2_CANAL SW: P43098"

/codon_start=1
/transl_table=1

/product="orf, hypothetical protein"

/protein_id="AAC74113.1"

/db_xref="GI:1787266"

/translation="MEDGVQAMRDYLAGLIDSPHEQVIANVYAKSEVPSIIKENIS LHLITHTVKMTESLDPTLNMPTVAFLISNKKYLGMLNDPAGVDQGRVWECRAFSB AKVFK"

275. 299 /gene="b1028"

/note="factor Sigma32; predicted +1 start at 1094983"

389. 1375 /gene="ycdU"

/note="synonym: b1029"

389. 1375 /gene="ycdU"

/function="orf; Unknown"

/note="O328; This 328 aa ORF is 19 pct identical (4 gaps) to 191 residues of an approx. 288 aa protein Y320_MYCB SW: P47562"

/codon_start=1
/transl_table=1

/product="orf, hypothetical protein"

/protein_id="AAC74114.1"

/db_xref="GI:1787267"

/translation="MIPDYLRTIRPOKRNLIYVAILGLIGFYWKNAFTFPEDEI LIMSAMSGLYQILPSYALFISLIGPVLFLFRMRTSYKQVAVSAVKKYK FLRSKRYAFVGRFLQEDIDFPSSNALSTFAKPLRLFLILIVIMMTLVSVL ATLVEMRIWEAFLLCYVCLIFYFCRPLMHNDFMACDMYRMGHFNK"

1253. 1280 /gene="ycdU"

/note="central position to predicted promoter: -104"

1253. 1280 /bound_moiety="Deor predicted site"

/gene="ycdU"

/note="No predicted promoter"

1337. 1365 /bound_moiety="Deor predicted site"

/gene="ycdU"

/note="factor Sigma70; predicted +1 start at 1096049"

1494. 1745 /gene="b1030"

/gene="b1030"
/function="orf; Unknown"
/note="O83; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_PABIT SW: P49065"

/codon_start=1
/transl_table=1

/product="orf, hypothetical protein"

/protein_id="AAC74115.1"

/db_xref="GI:1787268"

/translation="MGGRRCQYTKKKVLSYELFFKXGGGIDSLRSPGQPAHC VHGLSMTSPVNDPRSGLLIPVCAIHEKRTFVRL"

1513. 1926 /gene="ycdV"

/note="synonym: b1031"

1513. 1926 /gene="ycdV"

/function="putative structure; Not classified"

/note="O137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein RS37_YEAST SW: P05759"

/codon_start=1
/transl_table=1

/product="putative ribosomal protein"

/protein_id="AAC74116.1"

/db_xref="GI:1787269"

/translation="WCNIRKKSPIFRSSLSINAVRGALTRFAPAGSPPLTAVTPT SNMISPVPEPSGGLIPVCAIHEKRTFVRL"

complement(2111..2198)
/gene="serX"

/note="synonym: b1032"

complement(2111..2198)
/gene="serX"

/product="tRNA-Ser"

/note="anticodon: GGA; CG Site No. 17878"

2269. 2296 /function="tRNA; tRNA"

2269. 2296 /note="factor Sigma70; predicted +1 start at 1096980"

2338. 2365 /note="factor Sigma70; predicted +1 start at 1097049"

2335. 2383 /note="factor Sigma70; predicted +1 start at 1097067"

2339. 3370 /gene="ycdW"

/note="synonym: b1033"

2393. 3370 /gene="ycdW"

/function="putative enzyme; Not classified"

/note="O325; This 325 aa ORF is 32 pct identical (2 gaps) to 177 residues of an approx. 312 aa protein YPRA_CORGL SW: P45637"

/codon_start=1
/transl_table=1

/product="putative dehydrogenase"

/protein_id="AAC74117.1"

/db_xref="GI:1787270"

/translation="WRIRFAPRPSGSMIDIFHPPTDTOMTEALRKAIPQAVRAM LEDTMSGOMRQVAVSOVLNFRFRPDYRIGQNSHQPEVREDFGTIGAGVL GSKVAGSQVTFPLRCMSRTKSGMPVQSPAGEBSATLSGCRVILNLTPTPTPV GIINQOLEKPDGAYLNLARGVAVEDDLAIDSGKVGAMLVNFRNRPPLPSP LMQHPRTITPVAITRPAVAVEISRTIAQLEKGRVCGVDRARG"

3278. 3308 /gene="ycdW"

/note="factor Sigma70; predicted +1 start at 1097992"

3383. 3410 /note="factor Sigma70; predicted +1 start at 1098094"

/gene="b1030"
/function="orf; Unknown"
/note="O83; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_PABIT SW: P49065"

/codon_start=1
/transl_table=1

/product="orf, hypothetical protein"

/protein_id="AAC74115.1"

/db_xref="GI:1787268"

/translation="MGGRRCQYTKKKVLSYELFFKXGGGIDSLRSPGQPAHC VHGLSMTSPVNDPRSGLLIPVCAIHEKRTFVRL"

1513. 1926 /gene="ycdV"

/note="synonym: b1031"

1513. 1926 /gene="ycdV"

/function="putative structure; Not classified"

/note="O137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein RS37_YEAST SW: P05759"

/codon_start=1
/transl_table=1

/product="putative ribosomal protein"

/protein_id="AAC74116.1"

/db_xref="GI:1787269"

/translation="WCNIRKKSPIFRSSLSINAVRGALTRFAPAGSPPLTAVTPT SNMISPVPEPSGGLIPVCAIHEKRTFVRL"

complement(2111..2198)
/gene="serX"

/note="synonym: b1032"

complement(2111..2198)
/gene="serX"

/product="tRNA-Ser"

/note="anticodon: GGA; CG Site No. 17878"

2269. 2296 /function="tRNA; tRNA"

2269. 2296 /note="factor Sigma70; predicted +1 start at 1096980"

2338. 2365 /note="factor Sigma70; predicted +1 start at 1097049"

2335. 2383 /note="factor Sigma70; predicted +1 start at 1097067"

2339. 3370 /gene="ycdW"

/note="synonym: b1033"

2393. 3370 /gene="ycdW"

/function="putative enzyme; Not classified"

/note="O325; This 325 aa ORF is 32 pct identical (2 gaps) to 177 residues of an approx. 312 aa protein YPRA_CORGL SW: P45637"

/codon_start=1
/transl_table=1

/product="putative dehydrogenase"

/protein_id="AAC74117.1"

/db_xref="GI:1787270"

/translation="WRIRFAPRPSGSMIDIFHPPTDTOMTEALRKAIPQAVRAM LEDTMSGOMRQVAVSOVLNFRFRPDYRIGQNSHQPEVREDFGTIGAGVL GSKVAGSQVTFPLRCMSRTKSGMPVQSPAGEBSATLSGCRVILNLTPTPTPV GIINQOLEKPDGAYLNLARGVAVEDDLAIDSGKVGAMLVNFRNRPPLPSP LMQHPRTITPVAITRPAVAVEISRTIAQLEKGRVCGVDRARG"

3278. 3308 /gene="ycdW"

/note="factor Sigma70; predicted +1 start at 1097992"

3383. 3410 /note="factor Sigma70; predicted +1 start at 1098094"

/gene="b1030"
/function="orf; Unknown"
/note="O83; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_PABIT SW: P49065"

/codon_start=1
/transl_table=1

/product="orf, hypothetical protein"

/protein_id="AAC74115.1"

/db_xref="GI:1787268"

/translation="MGGRRCQYTKKKVLSYELFFKXGGGIDSLRSPGQPAHC VHGLSMTSPVNDPRSGLLIPVCAIHEKRTFVRL"

1513. 1926 /gene="ycdV"

/note="synonym: b1031"

1513. 1926 /gene="ycdV"

/function="putative structure; Not classified"

/note="O137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein RS37_YEAST SW: P05759"

/codon_start=1
/transl_table=1

/product="putative ribosomal protein"

/protein_id="AAC74116.1"

/db_xref="GI:1787269"

/translation="WCNIRKKSPIFRSSLSINAVRGALTRFAPAGSPPLTAVTPT SNMISPVPEPSGGLIPVCAIHEKRTFVRL"

complement(2111..2198)
/gene="serX"

/note="synonym: b1032"

complement(2111..2198)
/gene="serX"

/product="tRNA-Ser"

/note="anticodon: GGA; CG Site No. 17878"

2269. 2296 /function="tRNA; tRNA"

2269. 2296 /note="factor Sigma70; predicted +1 start at 1096980"

2338. 2365 /note="factor Sigma70; predicted +1 start at 1097049"

2335. 2383 /note="factor Sigma70; predicted +1 start at 1097067"

2339. 3370 /gene="ycdW"

/note="synonym: b1033"

2393. 3370 /gene="ycdW"

/function="putative enzyme; Not classified"

/note="O325; This 325 aa ORF is 32 pct identical (2 gaps) to 177 residues of an approx. 312 aa protein YPRA_CORGL SW: P45637"

/codon_start=1
/transl_table=1

/product="putative dehydrogenase"

/protein_id="AAC74117.1"

/db_xref="GI:1787270"

/translation="WRIRFAPRPSGSMIDIFHPPTDTOMTEALRKAIPQAVRAM LEDTMSGOMRQVAVSOVLNFRFRPDYRIGQNSHQPEVREDFGTIGAGVL GSKVAGSQVTFPLRCMSRTKSGMPVQSPAGEBSATLSGCRVILNLTPTPTPV GIINQOLEKPDGAYLNLARGVAVEDDLAIDSGKVGAMLVNFRNRPPLPSP LMQHPRTITPVAITRPAVAVEISRTIAQLEKGRVCGVDRARG"

3278. 3308 /gene="ycdW"

/note="factor Sigma70; predicted +1 start at 1097992"

3383. 3410 /note="factor Sigma70; predicted +1 start at 1098094"

/gene="b1030"
/function="orf; Unknown"
/note="O83; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_PABIT SW: P49065"

/codon_start=1
/transl_table=1

/product="orf, hypothetical protein"

/protein_id="AAC74115.1"

/db_xref="GI:1787268"

/translation="MGGRRCQYTKKKVLSYELFFKXGGGIDSLRSPGQPAHC VHGLSMTSPVNDPRSGLLIPVCAIHEKRTFVRL"

1513. 1926 /gene="ycdV"

/note="synonym: b1031"

1513. 1926 /gene="ycdV"

/function="putative structure; Not classified"

/note="O137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein RS37_YEAST SW: P05759"

/codon_start=1
/transl_table=1

/product="putative ribosomal protein"

/protein_id="AAC74116.1"

/db_xref="GI:1787269"

/translation="WCNIRKKSPIFRSSLSINAVRGALTRFAPAGSPPLTAVTPT SNMISPVPEPSGGLIPVCAIHEKRTFVRL"

complement(2111..2198)
/gene="serX"

/note="synonym: b1032"

complement(2111..2198)
/gene="serX"

/product="tRNA-Ser"

/note="anticodon: GGA; CG Site No. 17878"

2269. 2296 /function="tRNA; tRNA"

2269. 2296 /note="factor Sigma70; predicted +1 start at 1096980"

2338. 2365 /note="factor Sigma70; predicted +1 start at 1097049"

2335. 2383 /note="factor Sigma70; predicted +1 start at 1097067"

2339. 3370 /gene="ycdW"

/note="synonym: b1033"

2393. 3370 /gene="ycdW"

/function="putative enzyme; Not classified"

/note="O325; This 325 aa ORF is 32 pct identical (2 gaps) to 177 residues of an approx. 312 aa protein YPRA_CORGL SW: P45637"

/codon_start=1
/transl_table=1

/product="putative dehydrogenase"

/protein_id="AAC74117.1"

/db_xref="GI:1787270"

/translation="WRIRFAPRPSGSMIDIFHPPTDTOMTEALRKAIPQAVRAM LEDTMSGOMRQVAVSOVLNFRFRPDYRIGQNSHQPEVREDFGTIGAGVL GSKVAGSQVTFPLRCMSRTKSGMPVQSPAGEBSATLSGCRVILNLTPTPTPV GIINQOLEKPDGAYLNLARGVAVEDDLAIDSGKVGAMLVNFRNRPPLPSP LMQHPRTITPVAITRPAVAVEISRTIAQLEKGRVCGVDRARG"

3278. 3308 /gene="ycdW"

/note="factor Sigma70; predicted +1 start at 1097992"

3383. 3410 /note="factor Sigma70; predicted +1 start at 1098094"

/gene="b1030"
/function="orf; Unknown"
/note="O83; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_PABIT SW: P49065"

/codon_start=1
/transl_table=1

/product="orf, hypothetical protein"

/protein_id="AAC74115.1"

/db_xref="GI:1787268"

/translation="MGGRRCQYTKKKVLSYELFFKXGGGIDSLRSPGQPAHC VHGLSMTSPVNDPRSGLLIPVCAIHEKRTFVRL"

1513. 1926 /gene="ycdV"

/note="synonym: b1031"

1513. 1926 /gene="ycdV"

/function="putative structure; Not classified"

/note="O137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein RS37_YEAST SW: P05759"

/codon_start=1
/transl_table=1

/product="putative ribosomal protein"

/protein_id="AAC74116.1"

/db_xref="GI:1787269"

/translation="WCNIRK

Qy 1 ATGAACCTTTAAAGTGGACAGCATTCGACGAAATCGTAGTTTCGCGAGTCTTGCT 60
 Db 8993 ATGAACTTTTAAAGTAGACAGCATTCGACGAAATCGTAGTTTCGCGAGTCTTGCA 9052
 Qy 61 GCGCTCGTTCCACATGCGGCGCGCGGTATCATTAACGCGCGCGCAATGTTCCGC 120
 Db 9053 GGTGTGTCTCTCAGTACGCGCGCGCGGTAAACGCGGTGGCGGTAAATATAGCGC 9112
 Qy 121 CCGGACTCAACGTTGAGCATTTATGATGCGGTTCCGTTACGCTGCGCTCTTGCA 180
 Db 9113 CCAATTCGAGCGAATTTACAGATGACGCGGTGCGGTAACTGCACTTGCTTGCA 9172
 Qy 181 AGGACGCGCGTAAATCTGAAACGACATTAACCGAGCGGTATGATGACGCGCGCAT 240
 Db 9173 ACTATGCGCGTAACTGATCTGATCTATTAACCGAGCGCGGTAAATGATGATGAT 9232
 Qy 241 GTAGCGCGCGGTGCGGATATGATCTATTAACGATCTGATGATGATGATGATGAT 300
 Db 9233 GTTGTGAGCGGCTCAGATGACGATGATGATGATGATGATGATGATGATGATGAT 9292
 Qy 301 GCCACCATGACCGATGAGCACTTAAATCTCCATATTACTGTGCGCAATGCGCGGT 360
 Db 9293 GCTACTCTTGATCAGTGGAAACGCAAAATCTGAAATGACGTTAAACAGTTGCGTGT 9352
 Qy 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCGACGTAATGTCGTCAGTT 420
 Db 9353 GGCACGCGTGTGCAATGATGACGATGATGATGATGATGATGATGATGATGATGAT 9412
 Qy 421 GGTTTGGCAACAACGCGCGCTAACCGATTTAA 456
 Db 9413 GCGTTGGTAAACAACGCGCGCTCATCAGTAACTAA 9448

FEATURES

source
 Takayama, Ikoma, Nara 630-01, Japan
 (E-mail:hmori@tc.aist-nara.ac.jp, Tel:81-7437-2-5660,
 Fax:81-7437-2-5669)
 Collaboration Information:
 Project:
 The Japan E.coli genome DNA sequencing project
 Group:
 The Japan E.coli genome DNA sequencing group
 Members: (1995.4 - 1996.3)
 Alba,H., Baba,T., Fujita,K., Hayaashi,K., Honjo,A.,
 Horiuchi,T., Ikemoto,K., Inada,T., Isono,S.,
 Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
 Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
 Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
 Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
 Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
 Yamamoto,Y. and Yano,M.
 Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishori@nibb.ac.jp
 Information operator:
 Name: Hirokazu Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmori@tc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://baw3.aist-nara.ac.jp/
 location/Qualifiers
 1..15047
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:83333"
 /map="23..7-24.0 min"
 /clone="Kohara clone #231"
 /note="Nucleotide position 1097101-1112147 from the
 initiation site of *thiA* (0 min.). This clone is from
 Kohara lambda miniset library."
 <1..276
 /note="unnamed protein product; ORF_ID:0230#5"
 /codon_start=1
 /transl_table=11
 /protein_id="BAJ5822.1"
 /db_xref="GI:4062605"
 /translation="DLFSSBNALSTFFAKPLMLRLFILVIEVMITLIVSLATVE
 WRWFEAYFLCVCPLIYFPCRFMLHNDPMACDMYFRMGHFNK"
 1152..2090
 /note="ORF_ID:0231#1
 similar to SwissProt Accession Number P45637"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical 33.0 kD protein in prob-proA
 intergenic region."
 /protein_id="BAJ5823.1"
 /db_xref="GI:4062606"
 /translation="MDIIFHPTFDQWIEALRKALIPQAVRAWKSNDSDADVALY
 WHPPEMLAGRDLCAYFALGAVDSILSKIQAPHEMLNPVPLFRLEDTGMBQOMEX
 AIVQVHLWFRPDDYRIQONSSHQPLPEYHREDPTIGLAGVLSKVAQSLQWTF
 PLRCWSPRTSRMPGVOSFAGEELSAFLSCRVILNLPMTPEVGIINOQLLEKLPD
 GAYLVNARGVHYVEDLLAALDSGVKGMALVFNREPLPSPSLMQHRRVITTHV
 AATTPRAVEYISRTIOLIEKSERVCQVDRARGY"
 2145..2882
 /note="unnamed protein product; ORF_ID:0231#2"
 /codon_start=1
 /transl_table=11
 /protein_id="BAJ5824.1"
 /db_xref="GI:4062607"
 /translation="WTPVDLHMTVASTAVSTLSYDIAQAKGKIKLPAITDHGPM
 EDAFHHMFIMNRIWPRVVDGVLIGIEVNIKNVGEIDCSKMFPSLDLIIAGFHS
 PVPAPDKATQMDATIASGNVHIIISHGNPKRYEIDVAVAEAAKHQVAALEINS
 SPFHSRRKSEDNCREVAAVRADAGWVALQSDSHTAFTMGEPFECLEKILDAVDPPPR

RESULT 13
 D90741 15047 bp DNA linear BCT 25-DEC-2002
 LOCUS Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
 DEFINITION D90741 AB001340
 ACCESSION D90741.1 GI:1651509
 VERSION D90741.1 GI:1651509
 KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
 csgA; ycdB; cIs; nov; mdog.
 SOURCE Escherichia coli K12
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1
 Oshima,T., Alba,H., Baba,T., Fujita,K., Hayaashi,K., Honjo,A.,
 Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
 Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
 Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
 Nishio,Y., Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K.,
 Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
 A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map
 DNA Res. 3 (3), 137-155 (1996)
 JOURNAL MEDLINE 97061202
 PUBLISHED 8905232
 REFERENCE 2
 Alba,H., Baba,T., Fujita,K., Hayaashi,K., Honjo,A., Horiuchi,T.,
 Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
 Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
 Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
 Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
 Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
 Yamamoto,Y. and Yano,M.
 The systematic sequencing of the Escherichia coli genome in Japan
 Unpublished
 3 (bases 1 to 15047)
 MORI,H.
 TITLE Direct Submission
 AUTHORS Submitted (29-JUL-1996) Hirokazu Mori, NARA Institute of Science
 and Technology, Res. & Edu. Center for Genetic Info.; 8916-5

CDS

```
ILNVSPRLNLEBSRGMAPIAEFADL
2906..3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=1
/product="Hypothetical protein H11543"
/protein_id="BA035825.1"
/db_xref="GI:4062608"
/translation="MNESTLCNVLGSLVYROPDELVLVPLTIRREKLAANPLEQ
DELTPLQKSCDMTVASADYNALFIDECAPVPYSAWEGATREAVATFISEGMDL
RTMAPLTRDAISAMWDELEDSESE"
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"
/codon_start=1
/transl_table=1
/protein_id="BA035826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAFSAIRGNMILLSTAITTGILSGIKWVAVSLGL
SVNGLIGCTAYFACPOGGLKGLAISATLISGVWMAVITIGSLAHFHEILGYITG
LVAFMLCIQAKOLLSPVGFPIGACATPAQGGMKVLPSLGLIFGYAMNSGLM
LAARSAXTHRQEIKNRA"
complement(4117..4950)
/gene="cs9g"
complement(4117..4950)
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Cs9g precursor."
/protein_id="BA035827.1"
/db_xref="GI:4062610"
/translation="MORFLVAVMLISGLTAPPEKAPRTIMPRAQSYKDLTHLPA
NLNARKITIAQENGTVAIINRPILOSITANINVESTIGSNVSGGVQVLRG
IGADTQIQLQIVANLVVNVGETISVMSKTIISYVQAVFFITDYRLBE
VGTISNEPVMCLMSAIFETGVIFLINDIGDLMLOKAKERONDILVKYRHSVPPE
S"
complement(4977..5393)
/gene="cs9f"
complement(4977..5393)
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Cs9f precursor."
/protein_id="BA035828.1"
/db_xref="GI:4062611"
/translation="MRVKAHVLLMLISPLSAGTWTFOFRNPFGCNPNNCAFLINS
AQAKSYKDSYNDFGIETPSALDNFTQALISGLISININTEKFERMTNDYIV
DIANDQQLNVDRTKQSTIQTQVSGQNNSTD"
complement(5418..5807)
/gene="cs9g"
complement(5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Cs9g precursor."
/protein_id="BA035829.1"
/db_xref="GI:4062612"
/translation="MKRYLRVIAAEFTFAAGNLHAEVVPGLTDHTVSSIGHDFY
RAFSKDSYDNGNTINERSARWGSWITTVQDVIQTFPLPKRPFKEKTVFAL
IOTREAKNRQJNLSTGDLADEF"
complement(5812..6462)
/gene="cs9d"
```

CDS

```
complement(5812..6462)
/gene="cs9d"
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=1
/product="Probable cs9AB operon transcriptional regulatory
protein."
/protein_id="BA035830.1"
/db_xref="GI:4062613"
/translation="MFEVHSIHGHTLLITSSLOATPLLOHMQSIAITGKLNIQ
RSLDDISSGSIILLDMHEADKLIHWODTLRKNNTKILILNTPEDYPRDIENMP
HINGVFYSMEDQRRVNGOGVLRGCTQGLASYLITHSGNRYNTEBALTTRE
KEILNLRIGASNNELRSLFSENVTKHLNLPKLIAYNRTQAVSAMNDNLR"
7217..7672
/gene="cs9B"
7217..7672
```

gene

CDS

Query Match 56.1%; Score 256; DB 1; Length 15047;
Best Local Similarity 72.6%; Pred. No. 1,2e-62;
Matches 331; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY	1	ATGAACTTTAAAGTGCGACATCCAGCAATCGTAGTTCTGCGAGTCTGGCT	60
DB	7713	ATGAACTTTAAAGTGCGACATCCAGCAATCGTAGTTCTGCGAGTCTGGCT	60
QY	61	GGGTGCTTCCCAATGCGCGCGCGGTATCATTAACGCGCGCAATAGTCCGCG	120
DB	7773	GGGTGCTTCCCAATGCGCGCGCGGTATCATTAACGCGCGCAATAGTCCGCG	120
QY	121	CGGAGTCAGTGAAGCATTTACAGTTCGCTAACCGTGCCTCTGCA	180
DB	7833	CGAATTCAGTGAAGCATTTACAGTTCGCTAACCGTGCCTCTGCA	180
QY	181	AGCAATGCCGTAATCTGAAACGACATTCACAGCGGTTAGTACCGCCGAT	240
DB	7893	ACTATGCCGTAATCTGAAACGACATTCACAGCGGTTAGTACCGCCGAT	240
QY	241	GTAAGCCAGGCGGTAATAGTATCTAATTAAGTCACTGAGATGTTGAGAAAT	300
DB	7953	GTTGTCAGGCGTCAATGACAGCTCATGATCGACCAACGCGGTTGTAACGC	300
QY	301	GCCACATGACACAGTGAACCGTAAATCTCGATATTAAGTCCGCAATAGCGGCT	360
DB	8013	GCTACTCTGATGAGTGAACCGCAAAATCTGAAATGACGTTAAACAGTTGCGT	360
QY	361	AATTAACGCGCGCTGTTATATGACACCGCATCTGATTCAGCGTAAATGTCAGTT	420
DB	8073	GCGAACGCGTGTGACGTTGACCAAGCTGATCTCTCCGTCACAGTCACTGAGTT	420
QY	421	GGTTTGGCAACAGCGCGCTAACCATTTAA	456
DB	8133	GGCTTTGTAACAGCGCGCATCATCATTTAA	8168

gene
CDS

```
RESULT 14
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Bjorck L., Olsen A., Wikstrom M. and Herwald H.
AUTHORS
Peptides
TITLE Patent: WO 03064446-A 15 07-AUG-2003;
JOURNAL Hansa Medical Research Aktiebolag (SE)
FEATURES
source
1..456
```

/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..456
/note="unassigned protein product"

CDS

/codon_start=1
/transl_table=1
/protein_id="CAB83541.1"
/db_xref="GI:3910402"
/translation="MKLLKVEAIAIVFSGSALAGVVPQYGGGNGHGGGNSGPNSE
LNIYQGGNSALALQTDARNSDLTIQHGSGNADVGGSDDSSIDLTRGFGNSAT
LDQWNGKNSMTYKQFGGNGAADVQTSNSSVVTVQVFGFNATAHQY"

ORIGIN

Query Match 55.8%; Score 254.4; DB 6; Length 456;
Best Local Similarity 72.4%; Pred. No. 2.8e-62;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGATCGAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGATCGAGTTCTGGAGTCTTGCT 60
QY 61 GGCCTCGTTCCACATGGGCGCGCGGTATATACGCGCGCGCATATGTTCCGCG 120
DB 61 GGTGTGTTCTCTGATACGGCGCGGTATACCGCGGTATATGTTCCGCGCG 120
QY 121 CCGACCTCAACGTTGAGATTATGATGCGTTCCGTTACGCTGCGCTTCTGCA 180
DB 121 CCAAAATCTGAGCTGAACATTACAGTACGCGTGGCGGTATCTGCACTTCTGCA 180
QY 181 AGGATGCCGCGTAAATCTGAAGACCATTCACGAGCGGTTATGTTAAAGCGCGCAT 240
DB 181 ACTGATGCCGTAATCTGACTTGAATTAACCGATGCGCGGTATGTTGCGCAT 240
QY 241 GTAGCCAGGCGTGGATTAATGATTAATGATGATGATGATGATGATGATGAT 300
DB 241 GTTGTGATGCGCTCAGATGACAGCTCATGATCTGACCGACGTTGCGTACG 300
QY 301 GCCACATGACGATGGAACGCTTAAATCTCGATTTATGTCGCGCAATACGCGCGT 360
DB 301 GCTACTCTTGAATGATGGAACGCGCAAAATCTGAATGATGATGATGATGAT 360
QY 361 AATAAGCGCGGCTGTTAATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 GCGACGCGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTTTGGCAACAACGCGCGCGCTTAACCACTATTA 456
DB 421 GCGTTGGTAACAACGCGCGCGCTCATGATTA 456

RESULT 15
ECOCSSGA 648 bp DNA linear BCT 13-JUL-1993

DEFINITION Escherichia coli curlin subunit (csgA) gene, complete cds.
VERSION L049379.1 GI:290424
KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS Olsen, A., Arntqvist, A., Hammar, M., Sukupolvi, S. and Normark, S.
TITLE The RpoS sigma factor relieves H-NS-mediated transcriptional
repression of csgA, the subunit gene of fibronectin-binding curli
in Escherichia coli

JOURNAL Mol. Microbiol. 7 (4), 523-536 (1993)
MEDLINE 93211294
PUBMED 8459772

COMMENT On Jun 11, 1993 this sequence version replaced gi:145630.
Original source text: Escherichia coli (sub_strain W1110, strain
K-12) (Library: Kohara) DNA.

FEATURES

SOURCE

Location/Qualifiers
1..648
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W1110"
/db_xref="taxon:562"
/feature_11b="Kohara"

gene

83..538
/gene="csgA"

CDS

83..538
/gene="csgA"

/note="precursor protein"

/codon_start=1
/evidence=experimental

/transl_table=1
/product="curlin subunit"

/protein_id="AAA23616.1"
/db_xref="GI:290425"

/translation="MKLLKVEAIAIVFSGSALAGVVPQYGGGNGHGGGNSGPNSE
LNIYQGGNSALALQTDARNSDLTIQHGSGNADVGGSDDSSIDLTRGFGNSAT
LDQWNGKNSMTYKQFGGNGAADVQTSNSSVVTVQVFGFNATAHQY"

sig_peptide

83..142
/gene="csgA"

/evidence=experimental

mat_peptide

143..535
/gene="csgA"

/product="curlin subunit"

/evidence=experimental

ORIGIN

Query Match 55.8%; Score 254.4; DB 1; Length 648;
Best Local Similarity 72.4%; Pred. No. 2.9e-62;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGATCGAGTTCTGGAGTCTTGCT 60
DB 83 ATGAACCTTTTAAAGTGGCAGCATTCGACGATCGAGTTCTGGAGTCTTGCT 60
QY 61 GGCCTCGTTCCACATGGGCGCGGTATATACGCGCGCGCATATGTTCCGCG 120
DB 143 GGTGTGTTCTCTGATACGGCGCGGTATACCGCGGTATATGTTCCGCG 120
QY 121 CCGACCTCAACGTTGAGATTATGATGCGTTCCGTTACGCTGCGCTTCTGCA 180
DB 143 GGTGTGTTCTCTGATACGGCGCGGTATACCGCGGTATATGTTCCGCG 180
QY 181 AGGATGCCGCGTAAATCTGAAGACCATTCACGAGCGGTTATGTTAAAGCGCGCAT 240
DB 203 CCAAAATCTGAGCTGAACATTACAGTACGCGTGGCGGTATCTGCACTTCTGCA 240
QY 241 GTAGCCAGGCGTGGATTAATGATTAATGATGATGATGATGATGATGATGAT 300
DB 263 ACTGATGCCGTAATCTGACTTGAATTAACCGATGCGCGGTATGTTGCGCAT 300
QY 301 GCCACATGACGATGGAACGCTTAAATCTCGATTTATGTCGCGCAATACGCGCGT 360
DB 383 GCTACTCTTGAATGATGGAACGCGCAAAATCTGAATGATGATGATGATGAT 360
QY 361 AATAAGCGCGGCTGTTAATGATGATGATGATGATGATGATGATGATGAT 420
DB 443 GCGACGCGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTTTGGCAACAACGCGCGCGCTTAACCACTATTA 456
DB 503 GCGTTGGTAACAACGCGCGCGCTCATGATTA 456

Search completed: March 15, 2004, 22:49:37
Job time : 1972.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds

(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 1 atgaactttaaaagtcgac.....ccacgcctaacagctatataa 456

Sequence: IDENTITY_NUC
Gap 10.0 , Gapext 1.0Scoring table: IDENTITY_NUC
Gap 10.0 , Gapext 1.0

Searched: 3373663 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

```
1: Geneseq19908:.*
2: Geneseq19908:.*
3: Geneseq19908:.*
4: Geneseq19908:.*
5: Geneseq19908:.*
6: Geneseq19908:.*
7: Geneseq19908:.*
8: Geneseq19908:.*
9: Geneseq19908:.*
10: Geneseq19908:.*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	2	AAQ87467
2	456	100.0	456	2	AAQ87467
3	456	100.0	456	2	AAQ87467
4	456	100.0	456	2	AAQ87467
5	456	100.0	456	2	AAQ87467
6	456	100.0	456	2	AAQ87467
7	456	100.0	456	2	AAQ87467
8	456	100.0	456	2	AAQ87467
9	456	100.0	456	2	AAQ87467
10	456	100.0	456	2	AAQ87467
11	456	100.0	456	2	AAQ87467
12	456	100.0	456	2	AAQ87467
13	456	100.0	456	2	AAQ87467
14	456	100.0	456	2	AAQ87467
15	456	100.0	456	2	AAQ87467
16	456	100.0	456	2	AAQ87467
17	456	100.0	456	2	AAQ87467
18	456	100.0	456	2	AAQ87467
19	456	100.0	456	2	AAQ87467
20	456	100.0	456	2	AAQ87467
21	456	100.0	456	2	AAQ87467
22	456	100.0	456	2	AAQ87467
23	456	100.0	456	2	AAQ87467

24	42.4	9.3	100	7	ACD68808
25	40.8	8.9	100	7	ACD68809
26	35.4	7.8	3411	5	AA888526
27	35.4	7.8	3412	5	AA889144
28	34.6	7.6	456	3	AA664620
29	34.6	7.6	456	3	AA664620
30	34.6	7.6	456	3	AA664620
31	34.6	7.6	456	3	AA664620
32	34.6	7.6	456	3	AA664620
33	34.6	7.6	456	3	AA664620
34	34.6	7.6	456	3	AA664620
35	34.6	7.6	456	3	AA664620
36	34.6	7.6	456	3	AA664620
37	34.6	7.6	456	3	AA664620
38	34.6	7.6	456	3	AA664620
39	34.6	7.6	456	3	AA664620
40	34.6	7.6	456	3	AA664620
41	34.6	7.6	456	3	AA664620
42	34.6	7.6	456	3	AA664620
43	34.6	7.6	456	3	AA664620
44	34.6	7.6	456	3	AA664620
45	34.6	7.6	456	3	AA664620

ALIGNMENTS

RESULT 1	AAQ87467	standard; DNA; 456 BP.
ID	AAQ87467	
XX	AAQ87467	
AC	AAQ87467	
XX	AAQ87467	
DT	25-MAR-2003	(revised)
DT	26-JUN-1995	(first entry)
XX	Agfa sequence.	
DE	Agfa sequence.	
XX	Salmonella; Agfa; vaccine; genetic immunization; ds.	
KM	Salmonella; Agfa; vaccine; genetic immunization; ds.	
XX	Salmonella.	
OS	Salmonella.	
XX	Key	Location/Qualifiers
FT	CDS	1..454
FT		/*tag= a
FT		/note= "Agfa"
XX	MO9425598-A2.	
PD	10-NOV-1994.	
XX	26-APR-1994;	94WO-IB000207.
XX	26-APR-1993;	93US-00054452.
PR	(UVI-)	UNIV VICTORIA INNOVATION & DEV CORP.
PA	(KING/)	KING J.
XX	Kay WW,	Collinson SK, Clouthier SC, Doran JL,
PI	WPI: 1994-358275/44.	
DR	P-PSDB; AAR74625.	
XX	Eliciting an immune response to Salmonella - using attenuated Salmonella	
PT	strains, vector constructs, or compans. contg. fimbrial type proteins.	
XX	Disclousure; Fig 7B; 95pp; English.	
XX	The DNA encodes the Salmonella Agfa protein. The DNA and isolated	
CC	proteins are used in genetic immunization and vaccine compositions.	
CC	respectively, to elicit an immune response to Salmonella in animals (e.g.	
CC	food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN	
CC	field.)	

XX Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
SQ

Query Match 100.0%; Score 456; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.2e-118;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAATGCGACGATTCGACGCAATCGTAGTTCTGCGAGTCTTGAGT 60
DB 1 ATGAACCTTTAAATGCGACGATTCGACGCAATCGTAGTTCTGCGAGTCTTGAGT 60
QY 61 GCGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
DB 61 GCGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
QY 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACCGTGGCTTGGCTGGCA 180
DB 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACCGTGGCTTGGCTGGCA 180
QY 181 AGCGATGCCCGTAATCTGAACGACCAATACCGAGCGGTTATGTTAAACGGCGCGAT 240
DB 181 AGCGATGCCCGTAATCTGAACGACCAATACCGAGCGGTTATGTTAAACGGCGCGAT 240
QY 241 GTAGGCCAGGGTGGGATATAGTACTATTGAATCTGACTGATGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGGATATAGTACTATTGAATCTGACTGATGTTTCAGAAATAT 300
QY 301 GCCACCATGACACAGTGAACGCTAAATCTCGGATATATCTGTGGCCATACGGCGGT 360
DB 301 GCCACCATGACACAGTGAACGCTAAATCTCGGATATATCTGTGGCCATACGGCGGT 360
QY 361 AATAAGCCCGCGTGGTATATCAAGCCGATCTGATTCAGGCTATGTTGGTCAAGTT 420
DB 361 AATAAGCCCGCGTGGTATATCAAGCCGATCTGATTCAGGCTATGTTGGTCAAGTT 420
QY 421 GGTTCGGCAACACCGCAAGCTAACAGTATTTAA 456
DB 421 GGTTCGGCAACACCGCAAGCTAACAGTATTTAA 456

RESULT 2

AAT74142
ID AAT74142 standard; DNA; 456 BP.

AC AAT74142;
XX
DT 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)
XX
DE Salmonella enteritidis 27655-3b agfa gene.
XX
KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
OS Salmonella enteritidis.
XX
FH Key Location/Qualifiers
FT CDS 1..456
FT /*tag= a
FT /label= agfa_gene_fragment
FT /transl_except= (pos:367..369,aa:Pro)

US5635617-A.
XX
PN 03-JUN-1997.
XX
PD 26-APR-1994; 94US-00233788.
XX
PF 26-APR-1993; 93US-00054452.
XX
PR (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PA Collinson SK, Kay WW, Doran JL;
XX

DR MPI; 1997-30986/28.
DR P-PSDB; AAM23570.
XX

PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
PS Claim 1; Col 19-112; 85bp; English.

XX The present sequence represents an isolated agfa gene derived from
CC Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC family Enterobacteria. It can also be used to provide proteins and
CC antibodies which can be used for assays. The nucleic acid sequence can be
CC used to provide probes or primers which can specifically hybridize to
CC nucleic acid molecules from greater than 99% of Salmonella strains that
CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
XX correct PF field.)

SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.2e-118;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAATGCGACGATTCGACGCAATCGTAGTTCTGCGAGTCTTGAGT 60
DB 1 ATGAACCTTTAAATGCGACGATTCGACGCAATCGTAGTTCTGCGAGTCTTGAGT 60
QY 61 GCGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
DB 61 GCGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
QY 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACCGTGGCTTGGCTGGCA 180
DB 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACCGTGGCTTGGCTGGCA 180
QY 181 AGCGATGCCCGTAATCTGAACGACCAATACCGAGCGGTTATGTTAAACGGCGCGAT 240
DB 181 AGCGATGCCCGTAATCTGAACGACCAATACCGAGCGGTTATGTTAAACGGCGCGAT 240
QY 241 GTAGGCCAGGGTGGGATATAGTACTATTGAATCTGACTGATGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGGATATAGTACTATTGAATCTGACTGATGTTTCAGAAATAT 300
QY 301 GCCACCATGACACAGTGAACGCTAAATCTCGGATATATCTGTGGCCATACGGCGGT 360
DB 301 GCCACCATGACACAGTGAACGCTAAATCTCGGATATATCTGTGGCCATACGGCGGT 360
QY 361 AATAAGCCCGCGTGGTATATCAAGCCGATCTGATTCAGGCTATGTTGGTCAAGTT 420
DB 361 AATAAGCCCGCGTGGTATATCAAGCCGATCTGATTCAGGCTATGTTGGTCAAGTT 420
QY 421 GGTTCGGCAACACCGCAAGCTAACAGTATTTAA 456
DB 421 GGTTCGGCAACACCGCAAGCTAACAGTATTTAA 456

RESULT 3

AAC64617
ID AAC64617 standard; DNA; 456 BP.

AC AAC64617;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis Agfa DNA sequence SEQ ID NO:1.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbrin; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX

PN WO20060102-A2.
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 PF 05-APR-1999; 99US-0127888P.
 XX
 PR (UUVI-) UNIV VICTORIA.
 XX
 PA White AP, Doran JL, Collison SK, Kay MW;
 PI WPI; 2000-672631/65.
 XX P-PSDB; AAB36341.
 DR
 XX
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 134; 139pp; English.
 XX
 XX The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SO Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
 Query Match 100.0%; Score 456; DB 3; Length 456;
 Beot Local Similarity 100.0%; Pred. No. 8.2e-118;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGAACCTTTAAAGTGGAGCATTCGACAGATCGTAGTTTCGGAGTCTTGCT 60
 DB 1 ATGAACCTTTAAAGTGGAGCATTCGACAGATCGTAGTTTCGGAGTCTTGCT 60
 QY 61 GCGCGTCTTCCACAATGGGCGCGCGGTATCATACGCGCGCGCATATGTTCCGGC 120
 DB 61 GCGCGTCTTCCACAATGGGCGCGCGGTATCATACGCGCGCGCATATGTTCCGGC 120
 QY 121 CCGGACTTAAGCGTTGAGCATTTATCATAGTTCGGCTAACGCTGCGCTTCTTGCA 180
 DB 121 CCGGACTTAAGCGTTGAGCATTTATCATAGTTCGGCTAACGCTGCGCTTCTTGCA 180
 QY 121 CCGGACTTAAGCGTTGAGCATTTATCATAGTTCGGCTAACGCTGCGCTTCTTGCA 180
 DB 121 CCGGACTTAAGCGTTGAGCATTTATCATAGTTCGGCTAACGCTGCGCTTCTTGCA 180
 QY 181 AGCGATGCGCGTAATCGAATCGAATCCGAGACGGTATGTAAGCGCGCGCAT 240
 DB 181 AGCGATGCGCGTAATCGAATCGAATCCGAGACGGTATGTAAGCGCGCGCAT 240
 QY 241 GTAGGCGGAGGTGCGATATAGTACTATTAAGTCACTCAAGATGTTTCAGAAATAT 300
 DB 241 GTAGGCGGAGGTGCGATATAGTACTATTAAGTCACTCAAGATGTTTCAGAAATAT 300

QY 301 GCCACATCCGACGAGTGGAAAGCTTAAACCTCCGATATTACTGCGCCATACGCGGT 360
 DB 301 GCCACATCCGACGAGTGGAAAGCTTAAACCTCCGATATTACTGCGCCATACGCGGT 360
 QY 361 AATAAGCGCGCGTGGTTAATCAGACCGCATCTGATTCACGCTAATGTCGTCAGGTT 420
 DB 361 AATAAGCGCGCGTGGTTAATCAGACCGCATCTGATTCACGCTAATGTCGTCAGGTT 420
 QY 421 GGTTCGCAACAACGCGCGCTAACCGATTTAA 456
 DB 421 GGTTCGCAACAACGCGCGCTAACCGATTTAA 456
 RESULT 4
 AAC64626
 ID AAC64626 standard; DNA; 456 BP.
 AC AAC64626;
 AC
 DT 26-FEB-2001 (first entry)
 XX
 XX AgfA::PT3#5 DNA sequence SEQ ID NO:19.
 DE
 XX *Salmonella*: agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen; ds.
 XX
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO20060102-A2.
 XX
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 XX
 PI (UUVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay MW;
 XX WPI; 2000-672631/65.
 XX P-PSDB; AAB36350.
 DR
 DR
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 137; 139pp; English.
 XX
 XX The present invention describes a recombinant agfA gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 3; Length 456;
 Best Local Similarity 94.1%; Pred. No. 9.2e-124;
 Matches 425; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTCTTGCGAGTCTTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTCTTGCGAGTCTTGCTGCT 60
QY 61 GGGGTGTTCCCAATGGGCGCGCGCGGTAACTCAATACGGCGGCGCAATAGTCCGGC 120
DB 61 GGGGTGTTCCCAATGGGCGCGCGCGGTAACTCAATACGGCGGCGCAATAGTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTAGAGGTTCCGCTAACGCTGCTTGCGAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTAGAGGTTCCGCTAACGCTGCTTGCGAA 180
QY 181 AGCGATGCCCGTAAATCTGAACACGATTAACCCAGACGGTTATGGTAAACGGCGCGAT 240
DB 181 AGCGATGCCCGTAAATCTGAACACGATTAACCCAGACGGTTATGGTAAACGGCGCGAT 240
QY 241 GTAGCGCAGGTTGCGGATTAATAGTACTATTGAACTGATCTAGATGCTTCAAGAAATAT 300
DB 241 GTAGCGCAGGTTGCGGATTAATAGTACTATTGAACTGATCTAGATGCTTCAAGAAATAT 300
QY 301 GCCACATTCGACCGTGAACGCTAAAACTCGATATTACTGCGCCCAATACGGCGGT 360
DB 301 GCCACATTCGACCGTGAACGCTAAAACTCGATATTACTGCGCCCAATACGGCGGT 360
QY 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGAGTATGCTGCTGAGGTT 420
DB 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGAGTATGCTGCTGAGGTT 420
QY 421 GGTTTGGCAACAACCGCGCTAACGATATTA 456
DB 421 GGTTTGGCAACAACCGCGCTAACGATATTA 456

```

RESULT 5

ACG64625 ID AAC64625 standard; DNA; 456 BP.

AC AAG64625;
 XX
 DT 26-FEB-2001 (first entry)

DE Agfa: PT3#4 DNA sequence SEQ ID NO:17.

XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen; ds.

OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

PA White AP, Doran JL, Collison SK, Kay W;

XX
 DR WPI: 2000-672631/65.
 DR P-PSDB; AAB36349.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PS protein useful for eliciting immune response in animal.
 XX Disclosure, Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/7A) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (7) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 3; Length 456;
 Best Local Similarity 93.2%; Pred. No. 1.1e-121;
 Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTCTTGCGAGTCTTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTCTTGCGAGTCTTGCTGCT 60
QY 61 GGGGTGTTCCCAATGGGCGCGCGCGGTAACTCAATACGGCGGCGCAATAGTCCGGC 120
DB 61 GGGGTGTTCCCAATGGGCGCGCGCGGTAACTCAATACGGCGGCGCAATAGTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTAGAGGTTCCGCTAACGCTGCTTGCGAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTAGAGGTTCCGCTAACGCTGCTTGCGAA 180
QY 181 AGCGATGCCCGTAAATCTGAACACGATTAACCCAGACGGTTATGGTAAACGGCGCGAT 240
DB 181 AGCGATGCCCGTAAATCTGAACACGATTAACCCAGACGGTTATGGTAAACGGCGCGAT 240
QY 241 GTAGCGCAGGTTGCGGATTAATAGTACTATTGAACTGATCTAGATGCTTCAAGAAATAT 300
DB 241 GTAGCGCAGGTTGCGGATTAATAGTACTATTGAACTGATCTAGATGCTTCAAGAAATAT 300
QY 301 GCCACATTCGACCGTGAACGCTAAAACTCGATATTACTGCGCCCAATACGGCGGT 360
DB 301 GCCACATTCGACCGTGAACGCTAAAACTCGATATTACTGCGCCCAATACGGCGGT 360
QY 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGAGTATGCTGCTGAGGTT 420
DB 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGAGTATGCTGCTGAGGTT 420
QY 421 GGTTTGGCAACAACCGCGCTAACGATATTA 456
DB 421 GGTTTGGCAACAACCGCGCTAACGATATTA 456

```


Db 421 GGTTTGGCAACAGCCAGGCTAACAGATTAA 456

RESULT 6
AAC64628 standard; DNA; 456 BP.

XX ID 'AAC64628 standard; DNA; 456 BP.
XX AAC64628;
XX
XX 26-FEB-2001 (first entry)
XX
XX
XX Agfa::PT#7 DNA sequence SEQ ID NO:23.
XX
XX
XX Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmomella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36352.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX PT protein useful for eliciting immune response in animal.
XX
XX
XX Disclosure; Page 137; 139pp; English.

XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEFI/TA) nucleation depended
XX assembly system of strains of Salmomella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmomella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

XX
XX Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
Best Local Similarity 93.0%; Pred. No. 3.7e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGCGATTTGGCGAATTCGTAAGTTTCTGGCGATGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGCGATTTGGCGAATTCGTAAGTTTCTGGCGATGCTGCT 60
Qy GGGCGTGGTCCCAATGGGGGCGGGGGGTAATCAATCAAGCGGGCGCAATAGTTCCGGC 120
Db GGGCGTGGTCCCAATGGGGGCGGGGGGTAATCAATCAAGCGGGCGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACCGTCCGCTTCGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACCGTCCGCTTCGCA 180
Qy 181 AGCGATCGCCGTAATCTGAAACGACCACTTACCAAGCGGTTATGTAACGGCGCAT 240
Db 181 AGCGATCGCCGTAATCTGAAACGACCACTTACCAAGCGGTTATGTAACGGCGCAT 240
Qy 241 GTAGCGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGCGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Qy 301 GCCACATCGACCGATGGAACGCTAAACCTCCGATTTACTGCGGCAATACGGCGGT 360
Db 301 GCCACATCGACCGATGGAACGCTAAACCTCCGATTTACTGCGGCAATACGGCGGT 360
Qy 361 AATTAACCGCGGCTGTTTATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Db 361 AATTAACCGCGGCTGTTTATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Qy 421 GGTTTGGCAACAGCGCAGGCTAACAGATTAA 456
Db 421 GGTTTGGCAACAGCGCAGGCTAACAGATTAA 456

RESULT 7
AAC64622
ID AAC64622 standard; DNA; 456 BP.

XX AAC64622;
XX
XX 26-FEB-2001 (first entry)
XX
XX
XX Agfa::PT#1 DNA sequence SEQ ID NO:11.
XX
XX Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmomella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36346.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX PT protein useful for eliciting immune response in animal.
XX
XX
XX Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP, 121 A; 112 G; 118 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
 Best Local Similarity 93.0%; Pred. No. 3.7e-121;

Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
 QY 61 GGGCTGCTTCCACAAATGGGCGGCGGGTAAATCAATACGGCGCAATGTTCCGGC 120
 DB 61 GGGCTGCTTCCACAAATGGGCGGCGGGTAAATCAATACGGCGCAATGTTCCGGC 120
 QY 121 CCGGACTCAACGTTGACATTTATCATAGTACGGTTCGGTAAACGCTGCTGCTGCTGCT 180
 DB 121 CCGGACTCAACGTTGACATTTATCATAGTACGGTTCGGTAAACGCTGCTGCTGCTGCT 180
 QY 181 AGCGATGCCCGTAAATCTGAACGACCATTTACCCAGACGGGTATGTTAAGCGCGCGAT 240
 DB 181 AGCGATGCCCGTAAATCTGAACGACCATTTACCCAGACGGGTATGTTAAGCGCGCGAT 240
 QY 241 GTAGGCGAGGGTGGCGATTAATGTAATTAAGTGAATGTAATGTTTCAAAAATATAT 300
 DB 241 GTAGGCGAGGGTGGCGATTAATGTAATTAAGTGAATGTAATGTTTCAAAAATATAT 300
 QY 301 GCGACATCGACAGTGAACGCTTAAACCTCCGATTTACTGTCGGCCAAATGCGCGGT 360
 DB 301 GCGACATCGACAGTGAACGCTTAAACCTCCGATTTACTGTCGGCCAAATGCGCGGT 360
 QY 361 AATAAGCGCGGCGGTGTTATTCAGACCGCATCTGATTCAGCGCTAATGCGTCAAGT 420
 DB 361 AATAAGCGCGGCGGTGTTATTCAGACCGCATCTGATTCAGCGCTAATGCGTCAAGT 420
 QY 421 GGTGTTGGCAACAGCGCACGGCTAACCAAGTATTA 456
 DB 421 GGTGTTGGCAACAGCGCACGGCTAACCAAGTATTA 456

RESULT 8
 AAC64629 standard; DNA; 456 BP.

XX AAC64629;
 XX 26-FEB-2001 (first entry)
 XX

DE Agfa::PT3#8 DNA sequence SEQ ID NO:25.
 XX *Salmonella*: agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN WO20060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PP 05-APR-2000; 2000WO-CO00356.
 XX
 FR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/55.
 XX
 DR P-PSDB; AAB36353.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP, 114 A; 108 G; 123 G; 111 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
 Best Local Similarity 93.0%; Pred. No. 3.7e-121;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
 QY 61 GGGCTGCTTCCACAAATGGGCGGCGGGTAAATCAATACGGCGCAATGTTCCGGC 120
 DB 61 GGGCTGCTTCCACAAATGGGCGGCGGGTAAATCAATACGGCGCAATGTTCCGGC 120
 QY 121 CCGGACTCAACGTTGACATTTATCATAGTACGGTTCGGTAAACGCTGCTGCTGCT 180
 DB 121 CCGGACTCAACGTTGACATTTATCATAGTACGGTTCGGTAAACGCTGCTGCTGCT 180

DB 121 CCGAGCTCAACGTTGAGCATTTATGATGACGTTCCGCTAACGCTGCTTTATGATCAG 180
QY 101 AGCGATGCCCGCTTAATTCGAAACGACCATTAACCCAGAGCGGTTATGTTACCGCCCGAT 240
DB 101 CTGGTTACCCGCTGTTGTTATCCCATGAAATGCGACATGCGAGGTTATGTTAAACGCGCCGAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATGTACTATTGAACGACTCAGATAGTGGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATGTACTATTGAACGACTCAGATAGTGGTTTCAGAAATAT 300
QY 301 GCCACCATCGACAGTGAACCGCTAAACCTCCGATATTACTGTGCGCCCAATACGCGCGT 360
DB 301 GCCACCATCGACAGTGAACCGCTAAACCTCCGATATTACTGTGCGCCCAATACGCGCGT 360
QY 361 AATAACGCGCGCTGTTATATGACACCGCATTTGATCCAGCTATAGTGGCTGAGGTT 420
DB 361 AATAACGCGCGCTGTTATATGACACCGCATTTGATCCAGCTATAGTGGCTGAGGTT 420
QY 421 GGTGTTGGCAACACGCGGCTAACCGATTTAA 456
DB 421 GGTGTTGGCAACACGCGGCTAACCGATTTAA 456

RESULT 9
AAC64623

ID AAC64623 standard; DNA; 456 BP.

AAC64623;

26-FEB-2001 (first entry)

Agfa::PT3#2 DNA sequence SEQ ID NO:13.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
vaccine; immune response; immunogen; ds.

Salmonella enteritidis.

Escherichia coli.

Synthetic.

MO200060102-A2.

12-OCT-2000.

05-APR-2000; 2000WO-CA000356.

05-APR-1999; 99US-0127888P.

(UVVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay MW;

WPI; 2000-672631/65.

P-PSDB; AAB36347.

Discloure; Page 136; 139P; English.

CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SfP17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;

Best Local Similarity 93.0%; Pred. No. 3.7e-121;

Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGACATTCGACAGCAATCGATGTTCTGCGAGTCTGCT 60

DB 1 ATGAAACTTTTAAAGTGGACATTCGACAGCAATCGATGTTCTGCGAGTCTGCT 60

QY 61 GCGCTGTTCCACATGCGCGCGCGCGGTATATCAACGCGCGCAATATGTTCCGCG 120

DB 61 GCGCTGTTCCACATGCGCGCGCGCGGTATATCAACGCGCGCAATATGTTCCGCG 120

QY 121 CCGAGCTCAACGTTGAGCATTTATGATGACGTTCCGCTAACGCTGCTTTATGATCAG 180

DB 121 CCGAGCTCAACGTTGAGCATTTATGATGACGTTCCGCTAACGCTGCTTTATGATCAG 180

QY 181 AGCGATGCCCGCTTAATTCGAAACGACCATTAACCCAGAGCGGTTATGTTACGCGCCGAT 240

DB 181 AGCGATGCCCGCTTAATTCGAAACGACCATTAACCCAGAGCGGTTATGTTACGCGCCGAT 240

QY 241 GTAGGCCAGGGTGGCGATTAATGTACTATTGAACGACTCAGATAGTGGTTTCAGAAATAT 300

DB 241 GTAGGCCAGGGTGGCGATTAATGTACTATTGAACGACTCAGATAGTGGTTTCAGAAATAT 300

QY 301 GCCACCATCGACAGTGAACCGCTAAACCTCCGATATTACTGTGCGCCCAATACGCGCGT 360

DB 301 GCCACCATCGACAGTGAACCGCTAAACCTCCGATATTACTGTGCGCCCAATATGATCAG 360

QY 361 AATAACGCGCGCTGTTATATGACACCGCATTTGATCCAGCTATAGTGGCTGAGGTT 420

DB 361 CTGGTTACCCGCTGTTGTTATCCCATGAAATGCGACATGCAACGTTATGTTGCTGAGGTT 420

QY 421 GGTGTTGGCAACACGCGGCTAACCGATTTAA 456

DB 421 GGTGTTGGCAACACGCGGCTAACCGATTTAA 456

RESULT 10

AAC64624

ID AAC64624 standard; DNA; 456 BP.

AAC64624;

26-FEB-2001 (first entry)

Agfa::PT3#3 DNA sequence SEQ ID NO:15.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
vaccine; immune response; immunogen; ds.

Salmonella enteritidis.

Escherichia coli.

Synthetic.

MO200060102-A2.

PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 FA (UYVI-) UNIV VICTORIA.
 XX
 P1 White AP, Doran JL, Collison SK, Kay WW;
 DR WPI: 2000-672631/65.
 DR P-PSDB; AAB36348.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong CC immunogens, which may be important for directing an immune response CC against the inserted epitope, and hybrid fimbriae are easy and CC the exemplification of the present invention.

Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;
 Query Match 88.4%; Score 403.2; DB 3; Length 456;
 Best Local Similarity 92.8%; Pred. No. 1.2e-120;
 Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

1 ATGAATCTTTTAAAGTGGAGCATTCGAGCATTCGTACTTTCTGGAGTCTGCGCT 60
 1 ATGAATCTTTTAAAGTGGAGCATTCGAGCATTCGTACTTTCTGGAGTCTGCGCT 60
 61 GACGCTGTTTCAACATGGGGGGGGGGGGGCGGTAATCATTAACGCGCGCGCAATGTTCCGGC 120
 61 GCGCTTATATATCGCTGTTTACCCCGTGTGTTACCAATGAATGGCAATCATCATCCGGC 120
 121 CCGAGCTCAAGTGTAGCATTTATCAGTACGTTGCGTTCGTAACGCGCTGCTTCGCA 180
 121 CCGAGCTCAAGTGTAGCATTTATCAGTACGTTGCGTTCGTAACGCGCTGCTTCGCA 180
 181 AGGATGCGCGTGAATCTGAAAGACATTACCAAGAGGGGTTATGTAAGCGCGCGCAT 240
 181 AGGATGCGCGTGAATCTGAAAGACATTACCAAGAGGGGTTATGTAAGCGCGCGCAT 240
 241 GTAGGCGAGGGTGGGATATAGTACTATTTGAACGACGCGCAATGGTTTGAATAAT 300
 241 GTAGGCGAGGGTGGGATATAGTACTATTTGAACGACGCGCAATGGTTTGAATAAT 300
 301 GCCACCATGACACAGTGAACGCTTAAATTCGATTTACTGTGGCAATACGGGCT 360

DB 301 GCCACCATGACACAGTGAACGCTTAAATTCGATTTACTGTGGCAATACGGGCT 360
 QY 361 AATTAAGCGCGCGTGGTTAATCAGACGCATCTGATTCGAGCGTAATGGCGCTCAGTT 420
 DB 361 AATTAAGCGCGCGTGGTTAATCAGACGCATCTGATTCGAGCGTAATGGCGCTCAGTT 420
 QY 421 GGTTTGGCAACAACGCGCAGCGCTAACCATATTAA 456
 DB 421 GGTTTGGCAACAACGCGCAGCGCTAACCATATTAA 456

RESULT 11
 AAC64630
 ID AAC64630 standard; DNA; 456 BP.
 XX
 AC AAC64630;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa: PT#9 DNA sequence SEQ ID NO:27.
 XX
 KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 P1 White AP, Doran JL, Collison SK, Kay WW;
 DR WPI: 2000-672631/65.
 DR P-PSDB; AAB36354.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong CC immunogens, which may be important for directing an immune response

CC Against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;

Best Local Similarity 92.5%; Pred. No. 4.1e-120;

Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGATGTTCTGGAGTCTTGCT 60

Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGATGTTCTGGAGTCTTGCT 60

QY 61 GGCCTGTTCCAAATGGGGGGGGGGGGGATATCAATACGGCGCGCAATAGTTCCGGC 120

Db 61 GGCCTGTTCCAAATGGGGGGGGGGGGGATATCAATACGGCGCGCAATAGTTCCGGC 120

QY 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGCTAACGGTGGCTGCTGCA 180

Db 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGCTAACGGTGGCTGCTGCA 180

QY 181 AGCGATGCGCGGTAATCTGAACGACCATTAACGAGCGGTTATGTAACGGCGCGAT 240

Db 181 AGCGATGCGCGGTAATCTGAACGACCATTAACGAGCGGTTATGTAACGGCGCGAT 240

QY 241 GTAGCCAGGGTGGGATATAGTACTTATTAAGTACTGACTGCAATGTTTCAAAATAT 300

Db 241 GTAGCCAGGGTGGGATATAGTACTTATTAAGTACTGACTGCAATGTTTCAAAATAT 300

QY 301 GCCACATCGACGATGGAGCGTAAATCCGATTTATCTGCGGCAATACGGCGGT 360

Db 301 GCCACATCGACGATGGAGCGTAAATCCGATTTATCTGCGGCAATACGGCGGT 360

QY 361 AATTAAGCGCGGCTGTTATCAAGCGCATCTGATTCAGCGTATGTCGTCAGGTT 420

Db 361 AATTAAGCGCGGCTGTTATCAAGCGCATCTGATTCAGCGTATGTCGTCAGGTT 420

QY 421 GGTGTTGGCAACAGCGGCGGCTTAACAGCATCTGATTCAGCGTATGTCGTCAGGTT 420

Db 421 GGTGTTGGCAACAGCGGCGGCTTAACAGCATCTGATTCAGCGTATGTCGTCAGGTT 420

RESULT 12

AAC64627

ID AAC64627 standard; DNA; 456 BP.

AC AAC64627;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#6 DNA sequence SEQ ID NO:21.

KM Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;

XX vaccine; immune response; immunogen; ds.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PP 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UUVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR P-PSDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PS protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SFP1/TFP) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, GsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (1) is

CC useful for the expression of recombinant Agfa protein which is useful for

CC eliciting an immune response in an animal. In a fimbrial presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fimbria protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention

XX

SQ

Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;

Best Local Similarity 92.5%; Pred. No. 4.1e-120;

Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 DNA sequence SEQ ID NO:29.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN MO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX
DR WPI; 2000-672631/65.
XX
DR P-PSDB; AAB36355.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 87.4%; Score 398.4; DB 3; Length 456;
Best Local Similarity 92.1%; Pred. No. 4.5e-119;
Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

1 ATGAAACTTTTAAAGTGGCAGCATTCGCAGCATCGTAGTTCTTGGCAGTGGCTTGGCT 60

Db 1 ATGAAACTTTTAAAGTGGCAGCATTCGCAGCATCGTAGTTCTTGGCAGTGGCTTGGCT 60
QY 61 GCGGTGCTTCCACATATGGGGGGGGGGGGGCTAATCATATACGGGGGCAATAGTCCGCG 120
Db 61 GCGGTGCTTCCACATATGGGGGGGGGGGGGCTAATCATATACGGGGGCAATAGTCCGCG 120
QY 121 CCGAATCTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
Db 121 CCGAATCTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
QY 181 AGCATGCTCCGTAATATGTAACGACCATTTACCCAGAGCGTTATGTAACGCGCCGAT 240
Db 181 AGCATGCTCCGTAATATGTAACGACCATTTACCCAGAGCGTTATGTAACGCGCCGAT 240
QY 241 GTAGCCAGAGGTGCGATTAATAGTACTATTTGAACGACTCAGAAATGTTTCAAAATAT 300
Db 241 GTAGCCAGAGGTGCGATTAATAGTACTATTTGAACGACTCAGAAATGTTTCAAAATAT 300
QY 301 GCCACCATGACACAGTGGAAACGCTAATAAATCTCCGATATTAATCTGCGCCAAATACGCGCGT 360
Db 301 GCCACCATGATATAGTGGTTAATCCGCTGTTTACCAATGAATGCAATGACAGCGCGT 360
QY 361 AATTAACGCGCGCTGTTAATCAGACCGCATGTAATTCACCGTATATGTCGTCAGATT 420
Db 361 AATTAACGCGCGCTGTTAATCAGACCGCATGTAATTCACCGTATATGTCGTCAGATT 420
QY 421 GGTTTTGGGAACAAAGCCACGCGCTAACCAATTTAA 456
Db 421 GGTTTTGGGAACAAAGCCACGCGCTAACCAATTTAA 456

RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
XX AAQ73066;
XX
AC 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
DE
XX
XX Salmonella; Agfa; vaccine; genetic immunization; ds.
OS
XX Salmonella enteritidis.
FH Key location/Qualifiers
FT 1..359
FT /*tag= a
FT /note= "Agfa"
FT /note= "Agfa"
FT 37..60
FT /*tag= d
FT /note= "TAF5 primer (pair with TAF6)"
FT 52..69
FT /*tag= b
FT /note= "TAF3 primer (pair with TAF4)"
FT complement(103..129)
FT /*tag= e
FT /note= "TAF6 primer (pair with TAF5)"
FT complement(292..402)
FT /*tag= c
FT /note= "TAF4 primer (pair with TAF3)"
XX
XX MO9425598-A2.
PN
XX
XX 10-NOV-1994.
PD
XX
XX 26-APR-1994; 94WO-IB000207.
PR
XX
XX 26-APR-1993; 93US-00054452.
XX

PA (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
P1 Kay WW, Collinson SK, Clouthier SC, Doran JL,
XX WPI: 1994-358275/44.
DR P-PSDB; AAR62761.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compans. confg. fimbrial type proteins.
XX
PS Disclosure: Fig 7A; 95pp; English.
XX
CC The DNA encodes the Salmonella enteritidis 27655-3b TnpHox mutant strain
CC agfa gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to Salmonella in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
SO Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 72.8%; Score 331.8; DB 2; Length 361;
Best Local Similarity 99.4%; Pred. No. 1.9e-97;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 64 GTCGTTCCACAATGGGGCGGGCGGCGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 123
DB 1 GTCGTACCAACAGTGGGGCGGGCGGCGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 60

QY 124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGCGCTGCTCGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGCGCTGCTCGCAAGC 120

QY 184 GATCCCGGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGGCGCGCATGTA 243
DB 121 GATCCCGGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGGCGCGCATGTA 180

QY 244 GGCAGGGTGGGATTAATGTAATCTTGAATGACTGACTGCAATGTTTCAAAATATATGCC 303
DB 181 GGCAGGGTGGGATTAATGTAATCTTGAATGACTGACTGCAATGTTTCAAAATATATGCC 240

QY 304 ACCATGACGACGAGTGAACGCTAAACAACTCCGATTAATCTGCGGCAATACGGCGGTAT 363
DB 241 ACCATGACGACGAGTGAACGCTAAACAACTCCGATTAATCTGCGGCAATACGGCGGTAT 300

QY 364 AACGCCGGCTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCCGGCTGTTAATCAGACCGCATCTGATTC 335

RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
XX AAT74141;
XX AC
XX 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
XX Salmonella enteritidis 27655-3b TnpHox mutant agfa gene fragment.
KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
XX Salmonella enteritidis.
OS
XX
FH Key location/Qualifiers
FT CDS 1..360
FT primer_bind /tag= a
FT /label= agfa_gene_fragment
FT primer_bind 16..60
FT /tag= b
FT /label= Primer_TAF5

FT primer_bind 52..69
FT /tag= c
FT /label= Primer_TAF3
FT primer_bind complement(103..126)
FT /tag= d
FT /label= Primer_TAF6
FT primer_bind complement(1294..312)
FT /tag= e
FT /label= Primer_TAF4
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX
XX WPI: 1997-309886/28.
XX
XX P-PSDB; AAM23569.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX
XX Claim 1; Col 107-110; 85pp; English.
XX
XX The present sequence represents an isolated agfa gene fragment derived
XX from Salmonella enteritidis 27655-3b TnpHox mutant strain. The nucleic
XX acid can be used to provide diagnostic assays for Salmonella and/or
XX enteropathogenic bacteria of the family Enterobacteria. It can also be
XX used to provide proteins and antibodies which can be used for assays. The
XX nucleic acid sequence can be used to provide probes or primers which can
XX specifically hybridise to nucleic acid molecules from greater than 98% of
XX Salmonella strains that are pathogenic to warm-blooded animals relative
XX to nucleic acid molecules from virtually all other microbial organisms.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
SO Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 72.8%; Score 331.8; DB 2; Length 361;
Best Local Similarity 99.4%; Pred. No. 1.9e-97;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 64 GTCGTTCCACAATGGGGCGGGCGGCGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 123
DB 1 GTCGTACCAACAGTGGGGCGGGCGGCGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 60

QY 124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGCGCTGCTCGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGCGCTGCTCGCAAGC 120

QY 184 GATCCCGGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGGCGCGCATGTA 243
DB 121 GATCCCGGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGGCGCGCATGTA 180

QY 244 GGCAGGGTGGGATTAATGTAATCTTGAATGACTGACTGCAATGTTTCAAAATATATGCC 303
DB 181 GGCAGGGTGGGATTAATGTAATCTTGAATGACTGACTGCAATGTTTCAAAATATATGCC 240

QY 304 ACCATGACGACGAGTGAACGCTAAACAACTCCGATTAATCTGCGGCAATACGGCGGTAT 363
DB 241 ACCATGACGACGAGTGAACGCTAAACAACTCCGATTAATCTGCGGCAATACGGCGGTAT 300

QY 364 AACGCCGGCTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCCGGCTGTTAATCAGACCGCATCTGATTC 335

Search completed: March 15, 2004, 17:51:39

Thu Mar 18 12:27:54 2004

Job time : 254.684 secs

us-09-543-407-1.rng

Page 12

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 456
Sequence: 1 atgaactcttaaaagtgc.....ccacgctaccagctatataa 456

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	1	US-08-233-788A-58 Sequence 58, Appl
2	331.8	72.8	361	1	US-08-233-788A-56 Sequence 56, Appl
3	334	7.5	1344	3	US-09-120-927-1 Sequence 1, Appl
4	34	7.5	1344	4	US-09-431-614-5 Sequence 5, Appl
5	33.4	7.3	7766	4	US-09-125-619-3 Sequence 3, Appl
6	31.2	6.8	1491	6	5486473-3 Patent No. 5486473
7	31.2	6.8	2436	4	US-09-540-236-492 Sequence 492, Appl
8	31.2	6.8	65792	4	US-09-596-002-31 Sequence 31, Appl
9	30.8	6.8	2547	3	US-08-508-761B-1 Sequence 1, Appl
10	30.8	6.8	1664976	4	US-08-916-421B-1 Sequence 1, Appl
11	30.6	6.7	4403765	3	US-09-103-840A-2 Sequence 2, Appl
12	30.6	6.7	4411529	3	US-09-103-840A-1 Sequence 1, Appl
13	29.8	6.5	1008	4	US-09-252-991A-10904 Sequence 10904, A
14	29.8	6.5	1077	4	US-09-252-991A-10818 Sequence 10818, A
15	29.8	6.5	1521	4	US-09-252-991A-10865 Sequence 10865, A
16	29.6	6.5	972	4	US-09-328-352-2055 Sequence 2055, Ap
17	29.4	6.4	422	4	US-09-252-991A-7641 Sequence 7641, Ap
18	29.4	6.4	1252	4	US-09-561-756-29 Sequence 29, Appl
19	29.4	6.4	1252	4	US-09-227-721-29 Sequence 29, Appl
20	29.4	6.4	1252	4	US-09-954-697-29 Sequence 29, Appl
21	29.4	6.4	1939	3	US-08-961-527-310 Sequence 310, App
22	29.4	6.4	3494	3	US-09-139-802-200 Sequence 200, App
23	29.4	6.4	3494	4	US-09-659-786-200 Sequence 200, App
24	29.4	6.4	3494	4	US-09-023-655-1006 Sequence 1006, Ap
25	29.4	6.4	3660	4	US-09-252-991A-12269 Sequence 12269, A
26	29.4	6.4	536165	4	US-09-214-808-1 Sequence 1, Appl
27	29.2	6.4	2052	1	US-08-443-104-5 Sequence 5, Appl

C 28	29.2	6.4	2052	1	US-08-238-130-6 Sequence 6, Appl
C 29	29.2	6.4	2052	1	US-08-442-859-5 Sequence 5, Appl
C 30	29.2	6.4	2052	2	US-08-398-489-5 Sequence 5, Appl
C 31	29.2	6.4	2052	2	US-08-894-772-1 Sequence 1, Appl
C 32	29.2	6.4	2052	2	US-09-207-844-1 Sequence 1, Appl
C 33	29.2	6.4	2052	4	US-09-252-509-1 Sequence 1, Appl
C 34	29.2	6.4	2052	5	PCT-US95-0534-5 Sequence 5, Appl
C 35	29.2	6.4	4403765	3	US-09-103-840A-2 Sequence 2, Appl
C 36	29.2	6.4	4411529	3	US-09-103-840A-1 Sequence 1, Appl
C 37	29	6.4	534	4	US-09-489-039A-2195 Sequence 2195, Ap
C 38	29	6.4	1818	4	US-09-489-039A-2162 Sequence 54, Appl
C 39	29	6.4	13794	4	US-08-956-171E-54 Sequence 4506, Ap
C 40	28.8	6.3	654	4	US-09-252-991A-4506 Sequence 354, App
C 41	28.8	6.3	720	4	US-09-328-352-354 Sequence 2, Appl
C 42	28.8	6.3	1662	1	US-08-565-386-2 Sequence 1, Appl
C 43	28.8	6.3	4529	1	US-08-565-386-1 Sequence 1, Appl
C 44	28.8	6.3	8906	4	US-09-027-169-5 Sequence 5, Appl
C 45	28.4	6.2	1026	3	US-07-751-691B-24 Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58

Query Match 100.0%; Score 456; DB 1; Length 456;
Best Local Similarity 100.0%; Pred. No. 3e-149;
Matches 456; Conservative 0; Mismatches 0; Indels 0;

QY 1 ATGAACTTTAAAGTGGAGCATTCGAGCAATGTAGTTCTGGAGTGTCTGAGCT 60
Db 1 ATGAACTTTAAAGTGGAGCATTCGAGCAATGTAGTTCTGGAGTGTCTGAGCT 60
QY 61 GGGCGTGTTCACAAATGGGGCGGGCGGTAAATCAATACGGGGCGGCAATAGTTCCGGC 120
Db 61 GGGCGTGTTCACAAATGGGGCGGGCGGTAAATCAATACGGGGCGGCAATAGTTCCGGC 120
QY 121 CCGAGCTCAACGTTGAGCAATTTATCAATACGGTTCCGCTAAACGCTGCGCTGCTGCA 180
Db 121 CCGAGCTCAACGTTGAGCAATTTATCAATACGGTTCCGCTAAACGCTGCGCTGCTGCA 180
QY 181 AGCGATGCCCGTAAATCTGAAACGACCAATTAACCAAGCGGTTATGTTACGGCGCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACCAATTAACCAAGCGGTTATGTTACGGCGCGAT 240
QY 241 GTAGGCGAGGCGGTGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGGCGAGGCGGTGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACCAAGTGGAGCTAAATACTCCGATTTATCTGCGGCAATACGGCGGT 360
Db 301 GCCACCATGACCAAGTGGAGCTAAATACTCCGATTTATCTGCGGCAATACGGCGGT 360
QY 361 AATAACGCGCGGTGTTATCAAGCCGCAATCTGATTCAGCGTAAATGTTGCGTCAAGTT 420
Db 361 AATAACGCGCGGTGTTATCAAGCCGCAATCTGATTCAGCGTAAATGTTGCGTCAAGTT 420
QY 421 GGTTCGCAACACGCGCAAGCTAAACCAATTTAA 456
Db 421 GGTTCGCAACACGCGCAAGCTAAACCAATTTAA 456

RESULT 2

US-08-233-788A-56
Sequence 56, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043,403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEBANDERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-788A-56

Query Match 72.8%; Score 331.8; DB 1; Length 361;
Best Local Similarity 99.4%; Pred. No. 66-106;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 64 GTCTTCCAAATGGGGCGGGCGGTAAATCAATACGGGGCGGCAATAGTTCCGGCCCG 123
Db 1 GTCTTCCAAATGGGGCGGGCGGTAAATCAATACGGGGCGGCAATAGTTCCGGCCCG 60
QY 124 GACTCAACGTTGAGCAATTTATCAATACGGTTCCGCTAAACGCTGCGCTGCTGCAAGC 183
Db 61 GACTCAACGTTGAGCAATTTATCAATACGGTTCCGCTAAACGCTGCGCTGCTGCAAGC 120
QY 184 GATGCCCGTAAATCTGAAACGACCAATTAACCAAGCGGTTATGTTACGGCGCGATGTA 243
Db 121 GATGCCCGTAAATCTGAAACGACCAATTAACCAAGCGGTTATGTTACGGCGCGATGTA 180
QY 244 GGCAGGCGGTGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 303
Db 181 GGCAGGCGGTGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
QY 304 ACCATGACCAAGTGGAGCTAAATACTCCGATTTATCTGCGGCAATACGGCGGTAAAT 363
Db 241 ACCATGACCAAGTGGAGCTAAATACTCCGATTTATCTGCGGCAATACGGCGGTAAAT 300
QY 364 AACGCGCGGTGTTATCAAGCCGCAATCTGATTCAGCGTAAATGTTGCGTCAAGTT 398
Db 301 AACGCGCGGTGTTATCAAGCCGCAATCTGATTCAGCGTAAATGTTGCGTCAAGTT 335

RESULT 3

US-09-120-927-1/c
Sequence 1, Application US/09120927
Patent No. 6262018
GENERAL INFORMATION:
APPLICANT: Kim, Jihyun Francis
APPLICANT: Bear, Steven V.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120,927
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,108
FILING DATE: 06-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1581
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-120-927-1

Query Match 7.5%; Score 34; DB 3; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.12;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGCGCTGCTTCCACATGGGCGCGCGGTATCATTAACGGCGCGCAAT 111
|||
DB 209 GCTCCGGTTGCGGATTAACCTGATTTGGCGATGACAGTCTTAAACAGTTCCGCCAAT 150
|||
QY 112 AGTTCGCGCGCGGACCTCAACGTTGAGCATTTATCAGTACGGTTCGCTACGCGCTT 171
|||
DB 149 AATTGAGCCATTGCTCAATGCTTTGCCGATCGATGGGTTTGGCCCAACGCAAAATTT 90
|||

QY 172 GC 173
||
DB 89 GC 88

RESULT 4

US-09-431-614-5/C
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:

; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schending, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; FILE REFERENCE: 21829/41 (EBC-003)
; CURRENT APPLICATION NUMBER: US/09/431,614
; EARLIER FILING DATE: 1999-11-02
; EARLIER APPLICATION NUMBER: 60/107,243
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-431-614-5

Query Match 7.5%; Score 34; DB 4; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.12;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGCGCTGCTTCCACATGGGCGCGCGGTATCATTAACGGCGCGCAAT 111
|||
DB 209 GCTCCGGTTGCGGATTAACCTGATTTGGCGATGACAGTCTTAAACAGTTCCGCCAAT 150
|||
QY 112 AGTTCGCGCGCGGACCTCAACGTTGAGCATTTATCAGTACGGTTCGCTACGCGCTT 171
|||
DB 149 AATTGAGCCATTGCTCAATGCTTTGCCGATCGATGGGTTTGGCCCAACGCAAAATTT 90
|||

QY 172 GC 173
||
DB 89 GC 88

RESULT 5

US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG

; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
; FILE REFERENCE: USH:234
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 7.3%; Score 33.4; DB 4; Length 7766;
Best Local Similarity 47.4%; Pred. No. 0.51;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 157 GCTAACGCTGCGCTGCTGCTGCAAAAGCATGCCGTAATCTGAAGACCATTAACCG 216
|||
DB 4432 GTTACGGCTGCGCTGCTGCTGAGACAGATGAGAGAGAGCTGACAGGCTCAAAATCCG 4491
|||
QY 217 AGCGTTATGTTAAACGCGCGCATGTAGCGCAGGTCGATTAATAGTATTAATGAACTG 276
|||
DB 4492 ATTGCTCTCTATTTGGGAAGGGTAATGAGATGCTGCGATTTTGGTAAGATGAAATG 4551
|||
QY 277 ACTCAGATGCTTCAAGAAATTAATCCACATGCACGACGTAAGAACTCCGAT 336
|||
DB 4552 AAGAGCATGATGATGATGCTGCTGCTATTTGAGGGGATGCTTAAGATGAAG 4611
|||
QY 337 ATTACTGTGCGCAATCGCGCGTAATACG 367
|||
DB 4612 TTGCTGTGAAGTATGATGATGAGAAAG 4642
|||

RESULT 6

5486473-3/C
; Patent No. 5486473
; APPLICANT: FUKITA, HIROYUKI, YOSHIDA, IWAO, TAKAGI, MITSUO,
; MANABE, SADAOKI, KONOSUKE
; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,049
; FILING DATE: 09-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 809,255
; FILING DATE: 18-DEC-1991
; APPLICATION NUMBER: 279,685
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 932,419
; FILING DATE: 19-NOV-1986
; SEQ ID NO: 3
; LENGTH: 1491
5486473-3

Query Match 6.8%; Score 31.2; DB 6; Length 1491;
Best Local Similarity 50.0%; Pred. No. 1.2;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 178 CAAAGCATGCCGTAATCTGAAGACGACCATTAACGAGCGGTTATGTTAAGCGGCC 237
|||
DB 204 CAAGTACTACCAAGGCTAGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAAC 145
|||
QY 238 GATGTAGGCGAGGATGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 297
|||
DB 144 CATGTTCAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 85
|||

QY 298 AATGCCACATGACCATGAGCGTAAACCTCC 333
DB 84 ATCGCTTCCAGTACCATGATCAACCCATGATGCTCC 49

RESULT 7
US-09-540-236-492/c
Sequence 492, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
FILE REFERENCE: 2709.2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 492
LENGTH: 2436
TYPE: DNA
ORGANISM: M.catarrhalis
US-09-540-236-492

Query Match
Best Local Similarity 53.7%; Pred. No. 1.6;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

QY 209 TTACCCAGCGGTTATGTTAAGCGCGCGATGAGCGCGAGGTGGGTAATAGTACTA 268
DB 1511 TTACCCAGCATTCATCATATCTTAAAGCAAGTTCCGGCATGGTGTGAATTAAGC 1452
QY 269 TTGA---ACTGACTCAGATGTTTCAGAAATATGCAACCATGACAGTGGAAAGCTA 325
DB 1451 TGAACCTTCATATCCAAATGCTTCAATAATAAACCGATGATGCGCTCTGCACAAAGA 1392
QY 326 AAAACCTCGATATTAATCTGTGCGCAATACCGCGGTATATAAGCC 369
DB 1391 ACTGTCACTCATGACCATCAACCAAACTTGGTTGCAATTGCC 1348

RESULT 8
US-09-596-002-31/c
Sequence 31, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Patterson, Chandra
APPLICANT: Berg, Kim, L.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
FILE REFERENCE: PM-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 31
LENGTH: 65792
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte template ID No. 6632636 31
PUBLICATION INFORMATION:
US-09-596-002-31

Query Match
Best Local Similarity 6.8%; Score 31.2; DB 4; Length 65792;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

QY 209 TTACCCAGCGGTTATGTTAAGCGCGCGATGAGCGCGAGGTGGGTAATAGTACTA 268
DB 1511 TTACCCAGCATTCATCATATCTTAAAGCAAGTTCCGGCATGGTGTGAATTAAGC 1452

DB 6819 TTACCCATTCATATATCTTAAAGCAAGTTCCGGCATGGTGTGAATTAATAGC 6760
QY 269 TTGA---ACTGACTCAGATGTTTCAGAAATATGCAACCATGACAGTGGAAAGCTA 325
DB 6759 TGAACCTTCATATCCAAATGCTTCAATAATAAACCGATGATGCGCTCTGCACAAAGA 6700
QY 326 AAAACCTCGATATTAATCTGTGCGCAATACCGCGGTATATAAGCC 369
DB 6699 ACTGTCACTCATGACCATCAACCAAACTTGGTTGCAATTGCC 6656

RESULT 9
US-08-508-761B-1
Sequence 1, Application US/08508761B
Patent No. 6027920
GENERAL INFORMATION:
APPLICANT: Joliff, Gwennael
APPLICANT: Guyonvarch, Armel
APPLICANT: Pufflication, Relano
APPLICANT: Duchillon, Francis
APPLICANT: Renaud, Michel
TITLE OF INVENTION: System for Protein Expression and
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
STREET: 400 Seventh St. N.W.
CITY: Washington D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,761B
FILING DATE: 31-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09652
FILING DATE: 29-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09870
FILING DATE: 02-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,409
REFERENCE/DOCKET NUMBER: P58525NA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2547 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Corynebacterium Melassecola
FEATURE:
NAME/KEY: CDS
LOCATION: 420..2390
US-08-508-761B-1

Query Match
Best Local Similarity 6.8%; Score 30.8; DB 3; Length 2547;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 280 CAGATGTTTCAGAAATATGCGCACATGAGCGAGTGAACCTTAATAAATCCGATATT 339
DB 1511 TTACCCAGCATTCATCATATCTTAAAGCAAGTTCCGGCATGGTGTGAATTAAGC 1452

Db 876 CAGAGTGGTGGACCATTTAGACCAACATGTAGACGACTAGCCGATTAAGAGCCATT 935
Gy 340 ACTGTCGGCCATATACGGCGGTATATACCC 369
Db 936 GTTGCTCTCCCATTCGATCGTGGAGAGCTCC 965

RESULT 10
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Built et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729
FILE REFERENCE: P8275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1148948)..(1148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855339)..(855339)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature

```

; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1115681)..(1115681)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

```

Query Match
Best Local Similarity 6.8%; Score 30.8; DB 4; Length 1664976;
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 288 TTTCAGAAATATATGCGACCATGACAGGCTTAATTAACCTGATTTACTGCG 347
DB 360315 TTTCATTAATGTTGTCATGATGTCATATCAAGAAATTAATCTGTTTATTCAT 360374
QY 348 CCAATACCGCGGTAATTAAGCGCGGCTGTTATCAGACCGCATCTGATTCACGCGTAA 407
DB 360375 CAACGCTAAACAAAGAAAGCTTTATGATATGACTCTTTTAAATCAATTTAAAG 360434
QY 408 GGTGCGTACGCTGCTT 425
DB 360435 GATGCTTAGGTTTAT 360452

```

```

RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

```

```

Query Match
Best Local Similarity 6.7%; Score 30.6; DB 3; Length 4403765;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 45 TGGCAGTCTCTGCTGCGCTGCTTCACATAGGCGCGCGGCTAATCATACGCGG 104
DB 1357218 TTGCCAGCGCTGCTGACGCGCGCGCTCACACGCGCGCTGAGAGATCCCTTGA 1357159
QY 105 CGGCATATGTTCCGCGCCGAGCTCAACGTTGACATTATCATGATCGCTTCCGCTAACGC 164
DB 1357158 AAACACCGCATGCGCTTAACGAGCCACGATAGGTGTGCCATGTCTATCTGTGCGCG 1357099
QY 165 TGGCCTTGCTCTG 177
DB 1357098 CCCGACGTGCTG 1357086

```

```

RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match
Best Local Similarity 6.7%; Score 30.6; DB 3; Length 4411529;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 45 TGGCAGTCTCTGCTGCGCTGCTTCACATAGGCGCGCGGCTAATCATACGCGG 104
DB 1357750 TTGCCAGCGCTGCTGACGCGCGCGCTTACACGCGCGCTGAGCTCGTGA 1357691
QY 105 CGGCATATGTTCCGCGCCGAGCTCAACGTTGACATTATCATGATCGCTTCCGCTAACGC 164
DB 1357690 AAACACCGCATGCGCTTAACGAGCCACGATAGGTGTGCCATGTCTATCTGTGCGCG 1357631
QY 165 TGGCCTTGCTCTG 177
DB 1357630 CCCGACGTGCTG 1357618

```

```

RESULT 13
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; PRIOR FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 10904
;; LENGTH: 1008
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

Query Match 6.5%; Score 29.8; DB 4; Length 1008;
Best Local Similarity 54.0%; Pred. No. 3;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 41 TTCTGCGAGTCTCTGCGTGGCGCTTCCCAATAGGGCGGCGCGGTATCATACG 100
Db 141 TGTATGCTGTGCTGCGTACCGAGGCGCTTGGCGCGCGCGCGCTG 200
Oy 101 GCGGCGCAATAGTTCCGCGCCCGACTCAAGTTGAGCATTTATCAGTACG 153
Db 201 GCGGGAAGAGAGCGCGCGCGCTGCGGCGCTTGTATCAAGCAGAT 253

RESULT 14
US-09-252-991A-10818/c
;; Sequence 10818, Application US/09252991A
;; Patent No. 6551795
;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 10818
;; LENGTH: 1077
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10818

Query Match 6.5%; Score 29.8; DB 4; Length 1077;
Best Local Similarity 54.0%; Pred. No. 3.1;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 41 TTCTGCGAGTCTCTGCGTGGCGCTTCCCAATAGGGCGGCGGTATCATACG 100
Db 530 TGTATGCTGTGCTGCGTACCGAGGCGCTTGGCGCGCGCGCGCTG 471
Oy 101 GCGGCGCAATAGTTCCGCGCCCGACTCAAGTTGAGCATTTATCAGTACG 153
Db 470 GCGGGAAGAGAGCGCGCGCGCTGCGGCGCTTGTATCAAGCAGAT 418

RESULT 15
US-09-252-991A-10865
;; Sequence 10865, Application US/09252991A
;; Patent No. 6551795
;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136

;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 10865
;; LENGTH: 1521
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10865

Query Match 6.5%; Score 29.8; DB 4; Length 1521;
Best Local Similarity 54.0%; Pred. No. 3.7;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 41 TTCTGCGAGTCTCTGCGTGGCGCTTCCCAATAGGGCGGCGCGGTATCATACG 100
Db 170 TGTATGCTGTGCTGCGTACCGAGGCGCTTGGCGCGCGCGCGCTG 229
Oy 101 GCGGCGCAATAGTTCCGCGCCCGACTCAAGTTGAGCATTTATCAGTACG 153
Db 230 GCGGGAAGAGAGCGCGCGCGCTGCGGCGCTTGTATCAAGCAGAT 282

Search completed: March 16, 2004, 04:36:14
Job time : 70.6647 secs

THIS PAGE BLANK (USPTO)

Result	Score	Query	Match	Length	DB	ID	Description
C 1	34.2	7.5	6.16	12	US-10-424-559-88945		Sequence 88945, App1
C 2	34	7.5	1344	9	US-09-835-684-6		Sequence 6, App1
C 3	34	7.5	1344	9	US-09-880-371-6		Sequence 6, App1
C 4	34	7.5	1344	9	US-09-879-248-5		Sequence 5, App1
C 5	34	7.5	1344	14	US-10-010-350-6		Sequence 6, App1
C 6	34	7.5	1344	15	US-10-441-726-5		Sequence 5, App1
C 7	34	7.5	2886	9	US-09-801-368-131		Sequence 111, App
C 8	33.4	7.3	7766	12	US-10-222-566-3		Sequence 3, App1
C 9	33.4	7.3	7766	14	US-10-223-162-3		Sequence 3, App1
C 10	33.4	7.3	7766	14	US-10-142-024-3		Sequence 3, App1
C 11	33.2	7.3	1662	14	US-10-156-761-1483		Sequence 3483, App
C 12	33.2	7.3	9025608	14	US-10-156-761-1		Sequence 1, App1
C 13	32.8	7.2	1059	9	US-09-738-626-482		Sequence 482, App
C 14	32.8	7.2	1113	9	US-09-738-626-481		Sequence 481, App
C 15	32.8	7.2	3309400	9	US-09-738-626-1		Sequence 1, App1

16	33.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appl
17	33.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appl
18	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
19	32.2	7.1	4830	12	US-10-282-1224-27869	Sequence 27869, A
20	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
21	31.8	7.0	930	12	US-10-282-1224-20335	Sequence 20335, A
22	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
23	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
24	31.8	7.0	24081	14	US-10-132-134-131	Sequence 13, Appl
25	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appl
26	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appl
27	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, Appl
28	31.2	6.8	2394	12	US-10-282-1224-26837	Sequence 26837, A
29	31	6.8	525	15	US-09-818-995-32144	Sequence 32144, A
30	31	6.8	2751	15	US-10-398-221-3563	Sequence 3563, Ap
31	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
32	30.8	6.8	1575	10	US-09-873-3676-255	Sequence 255, Appl
33	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appl
34	30.6	6.7	1164	12	US-10-282-1224-13500	Sequence 13500, A
35	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appl
36	30.6	6.7	2301	9	US-09-815-242-9044	Sequence 4844, Ap
37	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 942, Ap
38	30.6	6.7	7521	14	US-10-080-170-646	Sequence 646, App
39	30.4	6.7	392	12	US-10-282-1224-25231	Sequence 25231, A
40	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
41	30.4	6.6	25220	11	US-09-584-429-364	Sequence 364, App
42	30.2	6.6	379	12	US-10-424-599-10524	Sequence 10524, A
43	30.2	6.6	654	12	US-10-425-114-28546	Sequence 28546, A
44	30.2	6.6	3203	12	US-10-282-1224-24039	Sequence 24039, A
45	30.2	6.6	3222	12	US-10-282-1224-6780	Sequence 6780, Ap

ALIGNMENTS

```

RESULT 1
US-10-424-599-88945/c
: Sequence 88945, Application US/10424599
: Publication No. US20040031072A1
: GENERAL INFORMATION:
: APPLICANT: La Rosa Thomas J
: APPLICANT: Kovalic David K
: APPLICANT: Zhou Yihua
: APPLICANT: Cao Yongwei
: TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53223)B
: CURRENT APPLICATION NUMBER: US/10/424,599
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 285684
: SEQ ID NO 88945
: LENGTH: 616
: TYPE: DNA
: ORGANISM: Glycine max
: FEATURE:
: NAME/KEY: unsure
: LOCATION: (1) ..(616)
: OTHER INFORMATION: unsure at all n locations
: FEATURE:
: OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
: US-10-424-599-88945

```

Query Match	7.5%	Score 34.2	DB 12	Length 616
Best Local Similarity	50.9%	Pred. No. 0.46		
Matches 81	Conservative 0	Mismatches 78	Indels 0	Gaps 0

Oy 10 GGGCGGCAATAGTTCGGGCCGGAGCTCAAGGTGAGATTTACATACGGTTCGGCTA 160
 Db 556 GGGCCCAAGGGTAACGCCCTTGAATCAAGCGCGGTCATGTGTGATGATGATGTTG 437
 Oy 161 AGCGTGGCTTGCTCTCGAAMCGATGCCCTTAATCTGAACGACCATTTACCGAGCG 220
 Db 496 ATGATGTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 437

QY 221 GTATGTACGCGCCGATGTAGCCAGGTCGGATA 259
DB 436 GTAATGTACTAGAGGTATGACGAGGAGGTTCCGGTA 398

RESULT 2
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US2002001937A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Qiu, Deyan
; APPLICANT: Remick, Dean
; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
; FILE REFERENCE: 21829/71
; CURRENT APPLICATION NUMBER: US/09/835,684
; PRIOR FILING DATE: 2001-04-16
; PRIOR APPLICATION NUMBER: 60/198,359
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match
Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGCGCTGCTTCACATGCGGCGCGGTAATCATTAACGCGCGCAAT 111
DB 209 GCTCCCGTGTGCGCATTTACTGATTTGCGCATAGCACTTAAACAGTTCCGCCAAT 150
QY 112 AGTTCGCGCGCGGACTCAAGCTTGAGCAATTATCAGTACGCTCCGCTAAGCGTGCCTT 171
DB 149 AATTGAGCCATTGCTCAATGTTGCCGATCGATGAGGTTGTTCCCGCAACGCAATTT 90
QY 172 GC 173
DB 89 GC 88

RESULT 3
US-09-880-371-6/c
; Sequence 6, Application US/09880371
; Patent No. US20020059658A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Derocher, Jay
; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
; FILE REFERENCE: 21829/91
; CURRENT APPLICATION NUMBER: US/09/880,371
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: 60/211,585
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match
Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGCGCTGCTTCACATGCGGCGCGCGGTAATCATTAACGCGCGCAAT 111
DB 209 GCTCCCGTGTGCGCATTTACTGATTTGCGCATAGCACTTAAACAGTTCCGCCAAT 150
QY 112 AGTTCGCGCGCGGACTCAAGCTTGAGCAATTATCAGTACGCTCCGCTAAGCGTGCCTT 171
DB 149 AATTGAGCCATTGCTCAATGTTGCCGATCGATGAGGTTGTTCCCGCAACGCAATTT 90
QY 172 GC 173
DB 89 GC 88

RESULT 4
US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match
Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGCGCTGCTTCACATGCGGCGCGCGGTAATCATTAACGCGCGCAAT 111
DB 209 GCTCCCGTGTGCGCATTTACTGATTTGCGCATAGCACTTAAACAGTTCCGCCAAT 150
QY 112 AGTTCGCGCGCGGACTCAAGCTTGAGCAATTATCAGTACGCTCCGCTAAGCGTGCCTT 171
DB 149 AATTGAGCCATTGCTCAATGTTGCCGATCGATGAGGTTGTTCCCGCAACGCAATTT 90
QY 172 GC 173
DB 89 GC 88

RESULT 5
US-10-010-390-6/c
; Sequence 6, Application US/10010390
; Publication No. US20030104979A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Leon, Ernesto
; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
; FILE REFERENCE: 21829/111
; CURRENT APPLICATION NUMBER: US/10/010,390
; PRIOR FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: 60/248,169
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora

157 GCTAACCGCTGGCTGCTCTGCAAAAGGATGCGCCCTTAATCTGAAACGACCATTAACCG 216
Best LocalSimilarity 47.4%; Pred. No. 3.3; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 111;

```

Db      4432 GTTACGCGTGGCGCTGCTGTGAGCAGATGAGAGAACCTGCGAGGCTACAAATCCG 4491
QY      217 AGCGGTTATGTTACGCGCCCGATGAGCCAGGTCGCGAATAATGTAATTAAGTCTG 276
DB      4492 ATTGCTGCTGCTATTGGAAGGGTATGAGATGAGTGGCGGATTGTTGTAAGATGAGT 4551
QY      277 ACTCAGATGTTGTTGAAATATATGCAACCATGACCACTGAGACGCTAAAACCTCCGAT 336
DB      4552 AAGAAAGATGATCAATGTTCTGCTGCTATTGTTGAGGGGGATGCTAAGGATGAAAG 4611
QY      337 ATTACTGTGCGCCAAATACGCGGTATATACG 367
DB      4612 TTTGCTGTGAAGATATGATGATGTGAGAAAG 4642

```

```

RESULT 9
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
; FILE REFERENCE: UTS# 234USD4
; CURRENT APPLICATION NUMBER: US/10/222,162
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/125,619
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-222-162-3

```

```

Query Match      7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY      157 GCTAACCTGGCGTTGCTGTCGAAAGCATGCCCCGTAATCTGAAGACCATTAACCCAG 216
DB      4432 GTTACGCGTGGCGCTGCTGTGAGCAGATGAGAGAACCTGCGAGGCTACAAATCCG 4491
QY      217 AGCGGTTATGTTACGCGCCCGATGAGCCAGGTCGCGAATAATGTAATTAAGTCTG 276
DB      4492 ATTGCTGCTGCTATTGGAAGGGTATGAGATGAGTGGCGGATTGTTGTAAGATGAGT 4551
QY      277 ACTCAGATGTTGTTGAAATATATGCAACCATGACCACTGAGACGCTAAAACCTCCGAT 336
DB      4552 AAGAAAGATGATCAATGTTCTGCTGCTATTGTTGAGGGGGATGCTAAGGATGAAAG 4611
QY      337 ATTACTGTGCGCCAAATACGCGGTATATACG 367
DB      4612 TTTGCTGTGAAGATATGATGATGTGAGAAAG 4642

```

```

RESULT 10
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.

```

```

; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
; FILE REFERENCE: UTS# 234USD1
; CURRENT APPLICATION NUMBER: US/10/143,024
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 09/125,619
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-143-024-3

```

```

Query Match      7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY      157 GCTAACCTGGCGTTGCTGTCGAAAGCATGCCCCGTAATCTGAAGACCATTAACCCAG 216
DB      4432 GTTACGCGTGGCGCTGCTGTGAGCAGATGAGAGAACCTGCGAGGCTACAAATCCG 4491
QY      217 AGCGGTTATGTTACGCGCCCGATGAGCCAGGTCGCGAATAATGTAATTAAGTCTG 276
DB      4492 ATTGCTGCTGCTATTGGAAGGGTATGAGATGAGTGGCGGATTGTTGTAAGGATGAGT 4551
QY      277 ACTCAGATGTTGTTGAAATATATGCAACCATGACCACTGAGACGCTAAAACCTCCGAT 336
DB      4552 AAGAAAGATGATCAATGTTCTGCTGCTATTGTTGAGGGGGATGCTAAGGATGAAAG 4611
QY      337 ATTACTGTGCGCCAAATACGCGGTATATACG 367
DB      4612 TTTGCTGTGAAGATATGATGATGTGAGAAAG 4642

```

```

RESULT 11
US-10-156-761-3483/C
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1662)
US-10-156-761-3483

```

```

Query Match      7.3%; Score 33.2; DB 14; Length 1662;

```

Best Local Similarity 52.1%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 50 GTGCTGCGGCGGCTGTTCCACAAATGGCGCGCGGTAAATCAATACGCGCGCGCA 109
Db 590 GCGGCGCGGCTGACGTCATACATCGCGCGGTTCATCGTCAACGCGCGCGCGC 531
Qy 110 ATAGTCCGCGCGGACTCAACGTTGAGCATTTACGATCGGTTCCGCTAACGCTGCGC 169
Db 530 TGGGCGCGCGCCATGCGCGCGAGTCCGGGTGAGGTGATCGCGCGCGAGGCGTCCG 471
Qy 170 TTGCTCTGCAAAAGCATGCCG 191
Db 470 TTGAACCTTCTTGGCGCGGACCG 449

RESULT 12
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.3%; Score 33.2; DB 14; Length 9025608;
Best Local Similarity 52.1%; Pred. No. 1.3e+02;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 50 GTGCTGCGGCGGCTGTTCCACAAATGGCGCGCGGTAAATCAATACGCGCGCGCA 109
Db 4331781 GCGGCGCGGCTGACGTCATACATCGCGCGGTTCATCGTCAACGCGCGCGC 4331840
Qy 110 ATAGTCCGCGCGGACTCAACGTTGAGCATTTACGATCGGTTCCGCTAACGCTGCGC 169
Db 4331841 TGGGCGCGCGCCATGCGCGCGAGTCCGGGTGAGGTGATCGCGCGAGGCGTCCG 4331900
Qy 170 TTGCTCTGCAAAAGCATGCCG 191
Db 4331901 TTGAACCTTCTTGGCGCGGACCG 4331922

RESULT 13
US-09-738-626-482
Sequence 482, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 482
LENGTH: 1059
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-482

Query Match 7.2%; Score 32.8; DB 9; Length 1059;
Best Local Similarity 47.2%; Pred. No. 1.9;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 184 GATGCCCGTAAATGTGAAGACCACTTACCCAGCGGTTATGTGAACGCGCGCATGTA 243
Db 76 GAGGACGTTTCCACCGGAGAGACCACTGCAAGCTCTTCCGTTAAGCATCCGATGCA 135
Qy 244 GCGGCGGCGGATTAAGTACTATTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 303
Db 136 GCGGCTAATAAAGTAACATCACCGTCTACACCTGAGCGTGAAGAAAGTGAAGTGAAG 195
Qy 304 ACCATGACAGGTGAAGCGCTAAACCTCGATTTACTGTCGCGCAATACGCGGTTAAT 363
Db 196 ATCAACAGGCGTTCAATGAAAGCAACCAACCAATATTAAGTTGAGGTGATCGCGCTGCT 255
Qy 364 AACGCGCGGCTGTTAATGACAGCGCATCTGA 395
Db 256 ACTGGCATCTGACTGCTCGCATGGAAGCTGA 287

RESULT 14
US-09-738-626-481/c
Sequence 481, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 481

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaacttttaaaagtgc.....ccacgctaccagctatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estbda:*
- 2: em_estbun:*
- 3: em_estcin:*
- 4: em_estcma:*
- 5: em_estcra:*
- 6: em_estcpl:*
- 7: em_estcro:*
- 8: em_estcra:*
- 9: gb_estc1:*
- 10: gb_estc2:*
- 11: gb_estc3:*
- 12: gb_estc4:*
- 13: gb_estc5:*
- 14: gb_estc6:*
- 15: em_estfun:*
- 16: em_estgom:*
- 17: em_estgom:*
- 18: em_estgom:*
- 19: em_estgom:*
- 20: em_estgom:*
- 21: em_estgom:*
- 22: em_estgom:*
- 23: em_estgom:*
- 24: em_estgom:*
- 25: em_estgom:*
- 26: em_estgom:*
- 27: em_estgom:*
- 28: gb_estc1:*
- 29: gb_estc2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	255	55.9	680	12	B0618688 B0618688
C 2	37.4	8.2	484	9	AV430994 AV430994
C 3	35.2	7.7	813	28	B2204853 B2204853
C 4	34.6	7.6	702	12	B0441624 B0441624

Result No.	Score	Query Match	Length	ID	Description
C 5	34.6	7.6	937	12	B0444255 GA_BA002
C 6	34.2	7.5	296	6	AV058630
C 7	34.2	7.5	523	12	B0334624
C 8	34.2	7.5	1036	12	B0671547
C 9	34	7.5	500	14	CA711477
C 10	34	7.5	500	14	CA711477
C 11	34	7.5	365	9	AA746477
C 12	33.8	7.4	558	14	CD374421
C 13	33.8	7.4	559	13	B0765156
C 14	33.8	7.4	559	13	B0765156
C 15	33.8	7.4	1098	29	CNS02GQP
C 16	33.8	7.4	1098	29	CNS02GQP
C 17	33.6	7.4	653	12	B0336964
C 18	33.6	7.4	653	12	B0336964
C 19	33.6	7.4	712	12	CF437982
C 20	33.4	7.3	505	6	AL809250
C 21	33.4	7.3	679	29	CC961526
C 22	33.4	7.3	695	28	BH663951
C 23	33.4	7.3	699	12	B1959328
C 24	33.4	7.3	700	28	BH651371
C 25	33.4	7.3	709	28	BH690914
C 26	33.4	7.3	715	13	B0444683
C 27	33.4	7.3	785	28	BH578793
C 28	33.4	7.3	815	28	BH461716
C 29	33.4	7.3	827	28	BH438448
C 30	33.4	7.3	830	28	BH715216
C 31	33.4	7.3	835	28	BH575194
C 32	33.4	7.3	853	28	BH578402
C 33	33.4	7.3	1026	28	B2458784
C 34	33.2	7.3	589	28	B2300068
C 35	33.2	7.3	625	10	BF501499
C 36	33.2	7.3	629	9	AU295325
C 37	33.2	7.3	788	9	AU139321
C 38	33.2	7.3	1646	10	BF579120
C 39	33	7.2	456	12	B0331339
C 40	33	7.2	524	12	B0367372
C 41	33	7.2	547	12	B0333958
C 42	33	7.2	553	12	B0336903
C 43	33	7.2	569	12	B0334047
C 44	33	7.2	574	12	B0330328
C 45	33	7.2	607	12	B0365766

ALIGNMENTS

RESULT 1
LOCUS B0618688/c
DEFINITION B0618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186D22 5', mRNA sequence.
ACCESSION B0618688
VERSION B0618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis

REFERENCE
AUTHORS Kohata, Y., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
CONTACT: Tadashi Shin-i
CENTER FOR GENETIC RESOURCE INFORMATION
NATIONAL INSTITUTE OF GENETICS
1111 Yata, Mishima, Shizuoka 411-8540, Japan
TEL: 81-559-81-6856
FAX: 81-559-81-6855
EMAIL: tshini@genes.nig.ac.jp
URL: The information of this clone is available through the following

FEATURES
source http://xenopus.nibb.ac.jp.
Location/Qualifiers
1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="X186b23"
/issue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochi normalized Xenopus early gastrula library"

ORIGIN

Query Match 55.9%; Score 255; DB 12; Length 680;
Best Local Similarity 72.4%; Pred. No. 3,86-62;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1 ATGAAACCTTTAAAGTCGACGATTCGACGATTCGAGTTCGCGAGTCTGCT 60
DB 593 ATGAAACCTTTAAAGTCGACGATTCGACGATTCGAGTTCGCGAGTCTGCA 534
QY 61 GGGGTGCTCCACANTGGCGCGCGCGGTATCTAAAGCGCGCGCAATAGTTCCGC 120
DB 533 GGTGTTGTTCTCAGACGCGCGCGGTATCTAAAGCGCGCGCAATAGTTCCGC 474
QY 121 CCGGACTCAACGCTGATTCATGATGAGGTTCCGCTAACGCGCGCTTGTCTGCA 180
DB 473 CCAAAATCTGAGCTGAACATTTACAGATGAGGTTCCGCTAACGCGCGCTTGTCTGCA 414
QY 181 AGCGATGCGCGCTTAACTTGAAACGCAATTCACGAGCGGTATGTTAACGCGCGCAT 240
DB 413 ACTGATGCGCGCTTAACTTGAAACGCAATTCACGAGCGGTATGTTAACGCGCGCAT 354
QY 241 GTAGCGCAGGTCGCGATATAGTATCTATGATGATGATGATGATGATGATGAT 300
DB 353 GTTGTGAGGCTGAGATGAGATGATGATGATGATGATGATGATGATGATGATGAT 294
QY 301 GCCACATCGACGAGTGAACGCTTAAACTCCGATATTCGTCGCGCAATAGCGCGT 360
DB 293 GCTACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234
QY 361 AATAACGCGCGCTGTTAATCAACGCGATTCATTCGACGATTCGAGTTCGAGTTC 420
DB 233 GCGAAGCGTGTGAGATTCACCACTGATTCATTCCTCCGTCACGAGTTCAGGTT 174
QY 421 GGTTCGCAACGCGCGCGCTTAACTCCGATATTCGTCGCGCAATAGCGCGT 456
DB 173 GGCCTTGATTAACGCGCGCTTAACTCCGATATTCGTCGCGCAATAGCGCGT 138

RESULT 2
AV430994 484 bp mRNA linear EST 23-AUG-2000
LOCUS AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION P027e10 r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
1 (bases 1 to 484)
Tabata, S., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
DNA Res. 7, 223-227 (2000)

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute

FEATURES
source Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers
1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="P027e10_r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 9,5;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 321 CGCTAAACCTCGATATTACTGCGCCATACGCGGTATTAAGCGCGGTGTTAA 380
DB 441 CCCAAACACAAAGTGTCCGCGCGCTCAAGACGCGCTCAATCGCTGCTGCGCAG 382
QY 381 TCAGACCGCATCTGATTCAGCGTATGTCGTCAGGTTGTTGCAACACGCA 439
DB 381 GCCAAAGCTGTGATCCAAAGTACTGTCGAGGTCGCGAAAGGCGTCCGCGCA 323

RESULT 3
BZ204853 813 bp DNA linear GSS 11-OCT-2002
LOCUS CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION CH230-390D21, genomic survey sequence.
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 813)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shyartbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, F., de Jong, P., and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 Mb01 segment
Unpublished (1999)
Other GSSs: CH230-390D21.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or erting_information.htm). BAC end
page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. .813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"

/note="Vector: pTRABAC1.3; Site 1: Mbol; Site 2: Mbol;
 CHORI-230 Rat (BN/Sevhad/MCM) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 813;
 Best Local Similarity 49.0%; Pred. No. 51;
 Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 178 CAAGCGATGCCGTTAATCTGAAACGACCATTCACAGCGGTTATGTAAACGGCGC 237
 Db 548 CAAGGACCTCCACATCAACACAGACACACTCAAACTAAGAGAAAAAATAGGAGCA 607
 Qy 238 GATGAGCCGAGGTGGCGATTAATAGTACTATTAAGCTGACGAAATGTTTCAGAAT 297
 Db 608 TCTGGAACATGGGCACTGGAATAAAATCTTGAACAAACCAATGGCTTGTGCT 667
 Qy 298 AATGCCACCATCGACGATGAGACGCTAAACTCCGATATTACTGTGCGCAATACGGC 357
 Db 668 AAGATCAAGATCGACCAATGGGATCTCATTAATGCAAGCTTGTGAGCAAGACAC 727
 Qy 358 GGTAAATACGCC 369
 Db 728 TGTGTTAGGAC 739

RESULT 4
 BG441624 702 bp mRNA linear EST 15-MAR-2001

LOCUS GA_Ea0014A14f Gossypium arboreum 7-10 dpa fiber library Gossypium
 accession BG441624 GI:13351276

ACCESSION BG441624.1 GI:13351276
 VERSION BG441624.1
 KEYWORDS EST.
 SOURCE Gossypium arboreum
 ORGANISM Gossypium arboreum

REFERENCE Wing, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 1 (bases 1 to 702)
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.

TITLE An integrated analysis of the genetics, development, and evolution
 of the cotton fiber

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu

Seq primer: TAATACGACTCACTATGAGG
 High quality sequence stop: 696.

FEATURES
 source location/Qualifiers

1..702
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultiivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0014A14f"
 /issue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 702;
 Best Local Similarity 46.5%; Pred. No. 71;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 181 AGGATGCCGTTAATCTGAAACGACCATTCACAGCGGTTATGTAAACGGCGCAT 240
 Db 185 AACATGACGGGAATGCGAATGGGAATGTTGGCAGCAGCAATCAGAGAACCCAT 244
 Qy 241 GTAGCCAGGATGCGGATTAATAGTACTATTGAATGACTGACAGATGTTTCAGAAT 300
 Db 245 GACATGTAATGAATGAAATGAGATGTTACTCAAAAGAAAAACATTCAGATTCAT 304
 Qy 301 GCCACCATCGACGATGAGAAAGCTAAAACTCCGATTTACTGTGCGCAATACGGCGT 360
 Db 305 GAAATGCTGACAGACCAACACCAACACATGAAAAACATTAATGAAAAATGCGGT 364
 Qy 361 AATAGCCCGGCGGTTAATGAGACCGCATCTGATTCAGCCGTAATGTCGTCAGGT 420
 Db 365 CAGACGGAGAAATGCGACTCAGATGACATGACATGAAATGTCGTCAGAT 424
 Qy 421 G 421
 Db 425 G 425

RESULT 5
 BG444255 937 bp mRNA linear EST 15-MAR-2001

LOCUS GA_Ea0023M14f Gossypium arboreum 7-10 dpa fiber library Gossypium
 accession BG444255 GI:13353907

ACCESSION BG444255.1 GI:13353907
 VERSION BG444255.1
 KEYWORDS EST.
 SOURCE Gossypium arboreum
 ORGANISM Gossypium arboreum

REFERENCE Wing, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 1 (bases 1 to 937)
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.

TITLE An integrated analysis of the genetics, development, and evolution
 of the cotton fiber

JOURNAL Unpublished (2000)

COMMENT Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu

Seq primer: TAATACGACTCACTATGAGG
 High quality sequence stop: 804.

FEATURES
 source location/Qualifiers

1..937
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultiivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0023M14f"
 /issue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 937;
 Best Local Similarity 46.5%; Pred. No. 81;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 181 AGGATGCCGTTAATCTGAAACGACCATTCACAGCGGTTATGTAAACGGCGCAT 240
 Db 185 AACATGACGGGAATGCGAATGGGAATGTTGGCAGCAGCAATCAGAGAACCCAT 244
 Qy 241 GTAGCCAGGATGCGGATTAATAGTACTATTGAATGACTGACAGATGTTTCAGAAT 300

Db 245 GACATGTAATGCAATGATGTTACTCAAAAACAAACGATTCAGTTCAAT 304
 QY 301 GCCACCATGACCAAGTGGAAAGCTTAAATTCCTGATATTAATGCGGCATACGGCGGT 360
 Db 305 GAAATGCTGCAAGAGCAACACCAACATGAAACATTAACATGAAATGCGGT 364
 QY 361 AATTAACCGCGCGCTGTTATCAGACCGCATCTGATTCAGGGTAAATGTCGTCAGTT 420
 Db 365 CAGGACCGGAGGAATCCACTCAGAGTAACTGACCAACATGAAATGTCGTCAGAT 424
 QY 421 G 421
 Db 425 G 425

RESULT 6
 AV058630 296 bp mRNA linear EST 23-JUN-1999
 LOCUS AV058630
 DEFINITION clone 1810054H02, mRNA sequence.
 ACCESSION AV058630
 VERSION AV058630.1 GI:5158377
 XREFS EST.
 KEYWORDS Mus musculus (house mouse)
 SOURCE Mus musculus
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Atawa, K., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nishitama, H., Oda, H., Sugahara, Y., Suzuki, H., Shimoto, Y., Shiraki, T., Sogabe, Y., Yoshino, M., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 RIKEN Mouse ESTs
 TITLE Unpublished (1999)
 JOURNAL Contact: Chile Owa
 COMMENT Genome Science Laboratory
 RIKEN

3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@rtc.riken.go.jp
 Thermotabilization and thermostabilization of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 source
 1. 296
 Location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1810054H02"
 /sex="male"
 /tissue_type="pancreas"
 /dev_stage="adult"
 /clone_1ib="Mus musculus pancreas C57BL/6J adult"

ORIGIN
 Query Match 7.5%; Score 34.2; DB 9; Length 296;
 Best local Similarity 53.3%; Pred. No. 63;
 Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 251 GTGGCGATATAGTACTATTAAGCATCAGCAATGCTTTAGAAATTAATGCAACCATCG 310
 Db 13 GTGGAGACCAAGTTTCTTAGCTACCAAGGCTGGTAAAGAAAGAGCATCAACTG 72

QY 311 ACCAGTGAACGCTTAAATCCGATATTAATGCTGCGCCCAATACGGCGTAAATACGCCG 370
 Db 73 TCCAGTGAAGCAAAATGTAAGAAAGAAATGATTATCCCAAGTTGAAATTAAGTGG 132
 QY 371 CGCTGTTAATCAAG 385
 Db 133 AGATGCCAAATATAGA 147

RESULT 7
 B0334624 523 bp mRNA linear EST 05-MAR-2002
 LOCUS B0334624
 DEFINITION B0334624 Dictyostelium discoideum cDNA library, Af dictyostelium
 ACCESSION B0334624
 VERSION B0334624.1 GI:19164754
 XREFS EST.
 KEYWORDS Dictyostelium discoideum
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum

REFERENCE
 AUTHORS Uzunihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tsunin@genes.nig.ac.jp

FEATURES
 source
 1. 523
 Location/Qualifiers
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dda47006"
 /sex="mat A"
 /dev_stage="Aggregation stage"
 /clone_1ib="Dictyostelium discoideum cDNA library, Af"

ORIGIN
 Query Match 7.5%; Score 34.2; DB 12; Length 523;
 Best local Similarity 57.1%; Pred. No. 82;
 Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 238 GATGTAGCGCAGGCTGCGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 297
 Db 398 GATGAAGATGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 457
 QY 298 AATGACCATGACCATGACCATGACCATGACCATGACCATGACCATGACCATGACCATG 342
 Db 458 AATTAACCAACCAACCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 502

RESULT 8
 BG671547 1036 bp mRNA linear EST 30-APR-2001
 LOCUS BG671547
 DEFINITION DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5',
 mRNA sequence.
 ACCESSION BG671547
 VERSION BG671547.1 GI:13893646
 XREFS EST.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus

REFERENCE
 AUTHORS Xiao, H.S., Huang, Q.H., Zhang, F.X., Bao, L., Lu, Y.J., Guo, C.,
 1 (bases 1 to 1036)

This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Gu Shouling Road, Zhangjiang Hi-Tech Park, Pudong New Area, P. R. China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chc.sh.cn)

PCR primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=NO.

```

source
1. .1036
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRNBUB05"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/clone_lib="Rat DRG library"

```

Qy	11	TAAGATGACAGCATTCGAGCAATGTAAGTTTCGAGAGTCCTGGCTGGCCTCTTC	70
Db	168	TGAAAGATACATGAGTGAACGTGTCGATTTGCAGACATTTGCTGTATGATAGCTCTGT	227
Qy	71	CACAAATGGGGGGGGGGGGTATCTATTAAGGGGGGGCAATGATTCGGGCCCCGCACTCA	138
Db	228	CTGTAGAGTCTCTCGTGGTAACGATGGCCAGAGTCAACATGATTCGTGGGGGGTTA	287
Qy	131	CGTTGAGCATTTATCATGTAAGGATTCGGCTCGCTTAACGCTGCAGCTTCTCTCGAAGCAGTCCC	190
Db	288	CATCTGAAGTGAAGAGTGGCTCTTTTGCACATCTGAATTTGATCTGTCAAGGAGACCTT	347

RESULT 9	BB581987/c	LOCUS	DEFINITION
BB581987	234 bp	mRNA	linear EST 30-NOV-2000
BB581987	RIKEN full-length enriched, adult male colon Mus musculus		
CDNA clone 9030005005	5', mRNA sequence.		

ORGANISM	REFERENCE
<i>Mus musculus</i>	1 (bases 1 to 234)
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus-	
1 (bases 1 to 234)	
Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T.,	

REFERENCE AUTHORS	1 (bases 1 to 234)
Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P., Hnagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T., Hodojama, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J.,	

The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel.: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
url: <http://genome.igsc.riken.go.jp/>
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,
Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermosensitization and thermooactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Iwata, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. *Genome Res.* 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. *Methods Enzymol.* 303,
13-44 (1999)
Please visit our web site (<http://genome.irc.riken.go.jp>) for
further details.

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="903005005"
/sex="male"
/tissue_type="colon"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="RRKEN full-length enriched, adult male colon"
/notes="Size 1 Salt 5' end 3' end, cDNA library was

```

/clone lib="RIKEN full-length enriched, adult male colon
 /note="Site 1: Sali; site 2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GACAGAGAGAGATCCAAAGCTCTTTTCTTTTCTTNN 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 10.0 and subtraction to Rot = 185.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GACAGAGAGATTCGACGATTAATTAATATATCCCCCCCC
 3']. cDNA was cloned into the XhoI and BamHI sites.
 Vector: a modified pluescript KS(+) after bulk excision
 from Lambda PLC I. Cloning sites, 5' end: Sali; 3' end:
 BamHI"

Query Match	7.5%	Score 34;	DB 10;	Length 234;
Best Local Similarity	51.3%	Pred. No. 65;		
Matches 79;	Conservative 0;	Mismatches 75;	Indels 0;	Gaps 0

Qy 35 TCGAGTTTCTGCAGTGCCTGGCTGCCTGTTCCACATGGGCGGCGGTATC 94
Db 200 TAGTTTGTAGTGCCTTATGCGGGTCTCTTCCACACAGGCGCGCGCTTCC 141

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-2

Perfect score: 456

Sequence: 1 atgaataaacatcgtatc.....ctgcgtcaccacgctaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pt: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vt: *
15: gb_ba: *
16: gb_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_ov: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_scs: *
28: em_un: *
29: em_vt: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_by: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	2067	1	SEU43280
2	452.8	99.3	5103	1	STAU2301
3	452.8	99.3	22411	1	AE008749
4	451.2	98.9	254050	1	AL627269
5	451.2	98.9	301983	1	AE016840
6	382.8	83.9	1048	1	STACRBA
7	348.4	76.5	2889	1	CSP515700
8	334.4	73.3	10370	1	AE015131
9	332.8	73.0	4680	1	ECCSGABDG
10	332.8	73.0	10346	1	AE000205
11	332.8	73.0	15047	1	D90741
12	332.8	73.0	282504	1	AE016981
13	332.8	73.0	306358	1	AE016759
14	331.2	72.6	1711	1	AE275733
15	331.2	72.6	10190	1	AE005315
16	331.2	72.6	327773	1	AP002554
17	313.4	68.7	2920	1	CPR515701
18	271	59.4	2883	1	ESAS15702
19	184	40.4	1212	1	EC0131756
20	47.8	10.5	10709	1	AE015532
21	47.4	10.4	301214	1	AE016786
22	44.2	9.7	11160	1	AE008209
23	44.2	9.1	11476	1	AE009405
24	41.4	9.1	42602	2	AC116971
25	41.1	9.0	7218	6	16494
26	40.6	8.9	62381	5	BX640594
27	40.4	8.9	168457	2	AC111425
28	40.4	8.9	178934	2	AC127966
29	40.2	8.8	193553	10	AC124196
30	39.8	8.7	295500	1	AP005954
31	38.2	8.4	298406	3	CEI758BA
32	37.8	8.3	174732	2	AL136161
33	37.8	8.3	191567	9	AL138702
34	37.6	8.2	2000	6	AX655393
35	37.4	8.2	120937	10	AL591911
36	37.4	8.2	225016	2	AC063967
37	37.2	8.2	42310	3	CBRG01B5
38	37	8.1	164504	2	AC123240
39	36.6	8.1	252420	3	AE014841
40	36.6	8.0	92378	5	BX537249
41	36.6	8.0	110000	2	AC095863_05
42	36.6	8.0	196383	2	AC113159
43	36.6	8.0	230058	2	AC108229
44	36.6	8.0	254497	2	AC119520
45	36.6	8.0	323930	1	AP003194

ALIGNMENTS

RESULT 1
LOCUS SEU43280
DEFINITION *Salmonella enteritidis* agfBAC operon: fimbrin-like protein precursor (agfB), thin aggregative fimbriae precursor (agfA), and AgfC (agfC) genes, complete cds.
ACCESSION U43280
VERSION U43280.1
KEYWORDS GI:1184712
SOURCE
ORGANISM *Salmonella enteritidis*
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.
REFERENCE
1 (bases 1 to 2067)
Doran, J.L., Collinson, S.K., Burian, J., Sarlos, G., Todd, E.C.,

Pred. No. is the number of results predicted by chance to have a

TITLE Munro, C.K., Kay, C.M., Baner, P.A., Peterkin, P.I. and Kay, W.W.
DNA-based diagnostic tests for *Salmonella* species targeting *agfa*,
the structural gene for thin, aggregative fimbriae
J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
JOURNAL 94013373
MEDLINE 8104955
PUBMED 2 (bases 1 to 2067)
REFERENCE Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and
AUTHORS Kay, W.W.
TITLE *Salmonella enteritidis agfBAC* operon encoding thin, aggregative
fimbriae
J. Bacteriol. 178 (3), 662-667 (1996)
JOURNAL 96146512
MEDLINE 8550497
PUBMED 3 (bases 1 to 2067)
REFERENCE Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
AUTHORS Direct Submission
TITLE Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
JOURNAL Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
Victoria, BC V8W 3P6, Canada
FEATURES Location/Qualifiers
source 1..2067
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrc"
571..576
598..603
696..1151
/gene="agfB"
CDS 696..1151
/gene="agfB"
/note="fimbria-like protein precursor"
/codon_start=1
/product="AgfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKNKLLFMWLTILGPGIATATNYDLARSEYFAVELSKSFN
QAATIQVGTDSNARVREGSKLSLVIQEGGNRAKVDGNTFYIISOTGNANDA
SISQAYGNSAAIIOKSGNKANITQYGTQTAIVVQKSHMALRYTOR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
CDS 753..1148
/gene="agfB"
/product="AgfB"
/note="putative signal sequence of 21 amino acids;
fimbria-like protein"
1193..1648
/gene="agfa"
1193..1648
CDS 1193..1648
/gene="agfa"
/note="thin aggregative fimbria subunit precursor; major
fimbrial subunit of thin aggregative fimbriae precursor"
/codon_start=1
/transl_table=1
/product="Agfa fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLTLYAAFAIIVSGSALAGVDPWGSGGNGHNGSGSPST
LSTIYGSANAALAIOSDARKSBTTTQSGYNGADVGADGADNSTITELTNGRRNAT
IDDMNAKNSDITTYGOTGNNALVNOTASDSVWVRQVGFNNATANY"
1292..1354
/gene="agfa"
1352..1645
/gene="agfa"
CDS 1352..1645
/gene="agfa"
/product="Agfa fimbria"
/note="thin aggregative fimbria subunit; major fimbrial
subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfc"
sig_peptide
mat_peptide
stem_loop
gene

CDS 1710..2036
/gene="agfc"
/function="unknown"
/codon_start=1
/transl_table=1
/product="Agfc"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MHTLLLAALNSQITFTTQGGDIYVLPVLTNPPCVQVIL
SVRGVGGSHYQKQTLSPANQPIELSRISVSSSDSVKIIYTVSDGSLHSQQ
WPESAQ"

ORIGIN

Query Match 100.0%; Score 456; DB 1; Length 2067;
Best Local Similarity 100.0%; Pred No. 3.8e-123;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAACAAATTTGTTATTTATGATTTGACATATCTGGGCGCTGGGATTGCAAC 60
DB 696 ATGAAAACAAATTTGTTATTTATGATTTGACATATCTGGGCGCTGGGATTGCAAC 755
QY 61 GCACAAATTTATGATCTGCTGCTGCTGACAAATTAATTTTGGGTAATGAATTAGCAAG 120
DB 756 GCACAAATTTATGATCTGCTGCTGCTGACAAATTAATTTTGGGTAATGAATTAGCAAG 815
QY 121 TCTTCATTTATACAGCGCGCCATTATTTGTCAGTCGCGACGATATATGTGCGACAGTA 180
DB 816 TCTTCATTTATACAGCGCGCCATTATTTGTCAGTCGCGACGATATATGTGCGACAGTA 875
QY 181 CGCCAGAGGATCAAACTATTTGCTGCTATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 240
DB 876 CGCCAGAGGATCAAACTATTTGCTGCTATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 935
QY 241 AAAGTCAGCAGAGGAGGATTAATTAATTTGCTGCTATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 300
DB 936 AAAGTCAGCAGAGGAGGATTAATTAATTTGCTGCTATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 995
QY 301 GATCCGATATATGCAAAAGCGCTTACGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 360
DB 996 GATCCGATATATGCAAAAGCGCTTACGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 420
QY 361 GAAATTAAGGCGCAATTTTCCAGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 480
DB 1056 GAAATTAAGGCGCAATTTTCCAGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTAATTTGTCAGTA 540
QY 421 CAGTCCGATATGCTATTTGCGCTACCCACGCTAA 456
DB 1116 CAGTCCGATATGCTATTTGCGCTACCCACGCTAA 1151

RESULT 2
STA2J301 5103 bp DNA linear BCT 15-NOV-2000
LOCUS *Salmonella typhimurium* cs9G, cs9F, cs9E, cs9D, cs9B, cs9A, and k
DEFINITION cs9G genes.
ACCESSION AJ002301
VERSION AJ002301.1 GI:2739232
KEYWORDS cs9A gene; cs9B gene; cs9C gene; cs9D gene; cs9E gene; cs9F gene;
SOURCE *Salmonella typhimurium*
ORGANISM *Salmonella typhimurium*
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; *Salmonella*.
REFERENCE 1 Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Normark, S.
AUTHORS Curli fibers are highly conserved between *Salmonella typhimurium*
TITLE and *Escherichia coli* with respect to operon structure and
JOURNAL regulation
MEDLINE J. Bacteriol. 180 (3), 722-731 (1998)
PUBMED 98117058
REFERENCE 2 (bases 1 to 5103)
AUTHORS Romling, U.

TITLE Direct Submission
JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES
Source
1. .5103
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="SR-11"
/db_xref="taxon:602"
CDS
<1..322
/note="ORF179a"
/codon_start=2
/transl_table=1
/product="hypothetical protein"
/protein_id="CA05311.1"
/db_xref="GI:2739233"
/db_xref="GOA:O54290"
/db_xref="SWISS-PROT:O54290"
/translation="CTLSGMVMAVLIIHGSALAPHLIEIVSLGTIVAFIMCIQAKO
LLISFPQTFIGACATFAGQDMRLVLPISLALGILFGAMNSGLMLASRRRHSANT
AVTK"
gene
complement(361..1194)
/gene="csbg"
CDS
complement(361..1194)
/gene="csbg"
/function="lipoprotein"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Csgb protein"
/protein_id="CA05312.1"
/db_xref="GI:2739234"
/db_xref="GOA:O54291"
/db_xref="SWISS-PROT:O54291"
/translation="MPRLILVAVLISGCTTAPKQAAKFTLMPRAQSYKDLTHLPA
PLNGKIPVSVYIDETGQFKPYSPASNSTAPVOSATMLVTALKDSHFIFLEQGHQ
NLNERRKIRAAQNGTGVAMNRIPLQSLTANIMVSGSIIGSVNYSGGVARYG
IGADTQYLDQIIVNLRVNVSTGEIILSVVTSKTIISYEVAQVPRPIQDRLBEE
IGTNSRPVMLCLMSALETGYVIFLINDIGDLNDLQNKADRONDLVKTRLSVPE
S"
gene
complement(1221..1637)
/gene="csgf"
CDS
complement(1221..1637)
/gene="csgf"
/codon_start=1
/transl_table=1
/product="Csgf protein"
/protein_id="CA05313.1"
/db_xref="GI:2739235"
/db_xref="SWISS-PROT:O54292"
/translation="MRVYHAYVLLMRSPLTWAGNMTFOFNPNGPNNNGSFLNLS
AQQNSYKDPAYNDPGLETPSALDNFTQALQSQILGGLTNINTGKRGKRVNTDPII
DIARDDQLQNLVDRKTRSTIEVSGLTQSTDF"
gene
complement(1664..2059)
/gene="csge"
CDS
complement(1664..2059)
/gene="csge"
/codon_start=1
/transl_table=1
/product="Csg protein"
/protein_id="CA05314.1"
/db_xref="GI:2739236"
/db_xref="SWISS-PROT:O54293"
/translation="MKRYLTWIVAELLFATGNLHANEVEVEGGLTDHTVSSIGHB
FYRAFSDKWESEYGLTINERPARGSGWTTITVNDVIFQFLFPMKRDFFKTVVF
ALAQTEBALNRQIDQLTSLSDLRDSE"
gene
complement(2064..2714)
/gene="csgd"
CDS
complement(2064..2714)
/gene="csgd"
/function="transcriptional regulator"
/codon_start=1
/transl_table=1

/product="Csgd protein"
/protein_id="CA05315.1"
/db_xref="GI:2739237"
/db_xref="GOA:O54294"
/db_xref="SWISS-PROT:O54294"
/translation="MFNIVASHSHGHTLLITRKPSLQATALLQHLKQSLATIGKLANIQ
RSLEDIADGCIIVMDMDEADKLIHYQDNLSRKNNNIKTLLNTPDQYPREIENMP
HINQVFATPDQSHVUSGLQGITRGCEYFQKLASYLITISGNVRYNSTESALLTHE
KELINKLRICASNNEIRSLPISENTYKTHLYNLFKLIAYAKRTQAVSMANDILRR"
gene
complement(2869..2896)
/gene="csgDERG"
-10_signal
complement(2869..2874)
/gene="csgDERG"
-35_signal
complement(2891..2896)
/gene="csgDERG"
gene
3344..3376
/gene="csgBA"
-35_signal
3344..3349
/gene="csgBA"
-10_signal
3371..3376
/gene="csgBA"
gene
3469..3924
/gene="csgb"
CDS
3469..3924
/gene="csgb"
/function="nucleator"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Csgb protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translation="MKRKLFPMLITLIGAGIATATVYDLARSEYFAVELSKSSFN
QAALIGQVGDNDARVYRQESKLSVTSQGGNRRARADQAGNVPAYIQTGNANDA
SISQASAGNSAALIIQKSGKNANITQYGTQKTVAVVQKQSHMAIRVQR"
sig_peptide
3469..3531
/gene="csbB"
mat_peptide
3532..3921
/gene="csbB"
/product="Csgb protein"
/function="nucleator"
gene
3966..4421
/gene="csGA"
CDS
3966..4421
/gene="csGA"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="CsgA protein"
/protein_id="CA05317.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/db_xref="SWISS-PROT:P55225"
/translation="MKLIKVAAPFAIIVSSALAGVYPWQGGGNNHGGNSGSPDT
LSIYQGSANAALQSDARKSETTITQSGYGNAGADVGGADNSTIELTQNGFNNAT
IDQWNAKNSDITGVQYGNNAALVNQTPADSSVMVROVFGNNATANQY"
sig_peptide
3966..4025
/gene="csGA"
mat_peptide
4026..4418
/gene="csGA"
/product="CsgA protein"
/function="fiber subunit"
4483..4809
/gene="csGC"
gene
4483..4809
/gene="csGC"
/codon_start=1
/transl_table=1
/product="CsgC protein"
/protein_id="CA05318.1"

```

/db_xref="GI:2739240"
/translation="MTTLLALASNTQITFTTGGDIYVIPOVTLNEPCVOYVLT
SYRDSGGGSHQQKQTLPLPANQPIELSRISVNISSDSXKIIVTSDQSILHSQ
WPSAQ"
sig_peptide
4483..4533
/gene="cs9c"
mat_peptide
4534..4806
/gene="cs9c"
/product="cs9c"
gene
4940..5103
/gene="ORF103"
4940..>5103
/gene="ORF103"
/codon_start=1
/transl_table=11
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:O54295"
/translation="MSVKKNPALGICAFIHSAGVQGVTHFGAIVEPL
CDISTHAENI"
ORIGIN
Query Match 99.3%; Score 452.8; DB 1; Length 5103;
Best Local Similarity 99.6%; Pred. No. 3.4e-122;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACCTGGTGGCTGGATTGCAACC 60
DB 3469 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACCTGGTGGCTGGATTGCAACC 60
QY 61 GCGACAAATATATATCTGGCTGCTTCAAGATATATTTTCCGTAATATGAAATTAAGCAAG 120
DB 3529 GCGACAAATATATATCTGGCTGCTTCAAGATATATTTTCCGTAATATGAAATTAAGCAAG 120
QY 121 TCTTCATTTAATCAGCGCGCATTTATGTCAGTCGCGACGATATATGTCGCAAGTA 180
DB 3589 TCTTCATTTAATCAGCGCGCATTTATGTCAGTCGCGACGATATATGTCGCAAGTA 180
QY 181 CGCAGGAGAGATCAAACTATTGTCGTTATTTTCAAGAAGAGCAAAATATATCGGCGC 240
DB 3649 CGCAGGAGAGATCAAACTATTGTCGTTATTTTCAAGAAGAGCAAAATATATCGGCGC 240
QY 241 AAAGTCGACGACGAGGAGATATTAATCTTGGCTATATTTGCAAAACGGGCAATGCCAAC 300
DB 3709 AAAGTCGACGACGAGGAGATATTAATCTTGGCTATATTTGCAAAACGGGCAATGCCAAC 300
QY 301 GATGCCAGTATATCGCAAGCGCTTACGTAATAGTCAGCTATTATCCAGAAAGGTTCT 360
DB 3769 GATGCCAGTATATCGCAAGCGCTTACGTAATAGTCAGCTATTATCCAGAAAGGTTCT 360
QY 361 GGAATAAGGCAATATATACCGATACGTCACGCAAAAACGAGTTGATGCAAGAA 420
DB 3829 GGAATAAGGCAATATATACCGATACGTCACGCAAAAACGAGTTGATGCAAGAA 420
QY 421 CAGTCGCATATGCTATTCGCGTCACCCCAACGCTTAA 456
DB 3889 CAGTCGCATATGCTATTCGCGTCACCCCAACGCTTAA 456

```

AUTHORS
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Portolillo, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2

TITLE
Nature 413 (6858), 852-856 (2001)

JOURNAL
MEDLINE
PUBMED
21534948
11677609
2 (bases 1 to 22411)

REFERENCE
The *Salmonella typhimurium* Genome Sequencing Project
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at BioCyc; <http://ecocyc.org/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and RegunDB; http://kinch.cfr.umax.mx:8950/db/regulondb_intro.frameset

FEATURES
source
This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m33 subclone.
Location/Qualifiers
1..22411

/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGGC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
434..1308
/gene="phoH"
/note="synonym: STM1126"
434..439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"
454..1308
/gene="phoH"
/note="similar to E. coli phoB-dependent, ATP-binding pho regulation component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=11
/product="PhoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCANRVQENAGGVYLTQKLSGRQKAVIKAREAKVLRDSS
RSHKOREBSYVLSVOMGVEALIGMARDSDTSVVKARNAOHVYNAITSKULIRAT
YDILKRLGASFMQYCLRPETIGKVEIPRAYMKRTFENAVILDEAQNVTAAQMKF
LTRIGENVTIVNDITQCDIPKVRSGLSDALEREDEDMVGIVHNNKDDCVRSALC
ORTIHAYS"
complement (1414..2303)

gene

```
CDS
/gene="STM127"
/complement(1414..2295)
/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
identity in aa 15 - 264"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEQPPRLKPKGLIDTLGAMQSLTPRASORIAQYTLAPROVQ
SSIALSRETQAGATVIRFCRTLGKGFOPKMDLIELATSDSSPLDVESE
SDDAIAGLKLONTISNVLSETLMDLMOQVLGVDDALRHCHSVYIGVSGITLAD
MKHLRMIGLEGDASNHNPMYMOATLILKADVAMGSHSTSPETHSLARQACA
TTVAITHNLGSSPLCEADPCLINGNRQGLGDSITGYKAOLFPVFDLLYTLVQSSPE
QARBSKLRITMALDMTK"
/complement(2298..2303)
/gene="STM127"
/note="putative RBS for STM127; RegulonDB:STMS1H001399"
/complement(2580..4088)
/gene="STM128"
/complement(2580..4076)
/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=1
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MTHTSRGIVNVYLVLFGLLAMLNGVYPSRQKTADYFRCGR
VPGAGSVFATLTSITFMSIPAKAFSDMTITIGYLAIALPLPFIYFIPFK
LKTSAEYLEARPDVRCRLFASMSFMLFHLGAIITFLVTLALRPFIADPVLV
LISVWCIIYTMGMEIGVITWDTIOGLISGSAIIFVICLKTQGGIDELFTVQQA
DKFPATQFHMSTESTEPVLMIGFLFANIOQFASODVORYIVTDSIEFTKTLT
NAKLAVIYVFFAIGSALFYVOHOPDLRAGNTGILLPLFVTEMPVGIADLIIA
AIFAAOSISSLNSISSCFNSDIYQRLSHKRTPRKMKIATLIVILVAGLSAAS
VLMVADSEIMDFNSLIGLNGGPMGLFMGLFFKRAANGSAVLGIIISVITVLGA
RYATDLNFFFYGVIGSLVISGVIAPLAPAPPLTLDEKPEKVTL"
/complement(4083..4088)
/gene="STM128"
/note="putative RBS for STM128; RegulonDB:STMS1H001400"
/complement(4413..5099)
/gene="STM129"
/complement(4413..5093)
/gene="STM129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSILARLEQSVHENGCIIVSCOPYPSPMKREIVAAQAQAA
AGFAVRIEGIEIENIVRPHLSVPIIGIKDLGSPRITPVYQDDVALAQADII
AFDAFSRSPVDISLTRLRIHGLLAMDSTYNEGISCHQKIEIGITLGSYTB
ITPVEPLAMVQLSHAGCRVIAEGRVYTPALANAIEHGAMAVTVGSATIRHICQ
WFSHAVKR"
/complement(5094..5099)
/gene="STM129"
/note="putative RBS for STM129; RegulonDB:STMS1H001401"
5599..6759
/gene="STM130"
5599..6759
/gene="STM130"
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=1
```

```
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MEQMKRPFKKMTLMLCFEVAITTSQVATLPDIPPLKNGTG
AIDNNGVIYVGLGTAQTSWKYKIDLKQKHMERIKSPGGAIBQSVSVFLNDELIVFG
GVGKNSPFLQVSDYQYKSPVKNQKDTISPVGLGHTGKALNEMTWLITGVN
EHIIDKXFIIDIAAADADESEKNKYIVYFPAADYQFPNKVIFLYNKENWKNAGEL
PGAGTASSSSVMGNKPLMLINGELKPELRDVIYRAMMDNKLTLNKSQPLPSRBEQ
QDEGLAGFSGYSHGVLLVGGGANFPAKQNTYNGKTSHEGINKKMRDEVYGLNGH
WQYGMKQKQPLGVSYSYGDVFLIGENAKKQVSVSYTSFTWRDNLILIK"
6793..7497
/gene="STM131"
6793..6798
/gene="STM131"
/note="putative RBS for STM131; RegulonDB:STMS1H001402"
6805..7497
/gene="STM131"
/note="similar to E. coli orf, hypothetical protein
(AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
identity in aa 4 - 239"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein"
/protein_id="AAL20062.1"
/db_xref="GI:16419647"
/translation="MKINKYLGMVSFIASSYVQAATLDYRHEADRTIRNDRIAI
IEKLPNGIGFYVDASVKSQVSDGSDHLDLVAALIBLGVSTNYKTDNFPVLPQFLI
FESGPDTSIYKPYLRQGNFDSGVYMAGRYRDYARVYASDDEKTRPDYIYGVYF
DEKLBNFTFMDSDQIKFDNKKTYNEHNVALMKLNSFTPYVEGVNAVVRNNTDER
QTRRYVLQYHF"
7768..9060
/gene="STM132"
7768..7773
/gene="STM132"
/note="putative RBS for STM132; RegulonDB:STMS1H001403"
7780..9060
/gene="STM132"
/note="similar to E. coli putative transport protein
(AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
identity in aa 25 - 418"
/codon_start=1
/transl_table=1
/product="putative sugar transport protein"
Query Match 99.3%; Score 452.8; DB 1; Length 22411;
Best Local Similarity 99.6%; Pred. No. 3.6e-122;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGAAAAAATAATTGATTATGATGTTGACAAATACGGGTCGCTGGGATTGCAAC 60
Db 17272 ATGAAAAAATAATTGATTATGATGTTGACAAATACGGGTCGCTGGGATTGCAAC 17331
QY 61 GCGCAAAATATGATCTGCTCGTTCAAAATATATTTTGGTAAATGAATTAAGCAAG 120
Db 17332 GCGCAAAATATGATCTGCTCGTTCAAGATATATTTTGGGTAATTAAGCAAG 17391
QY 121 TCTTCATTATACGGGGCCCATTTGTCACAGTCGGCACGGTAATTAAGTCCAGAGTA 180
Db 17392 TCTTCATTATACGGGGCCCATTTGTCACAGTCGGCACGGTAATTAAGTCCAGAGTA 17451
QY 181 CGGCAGGAAGATCAAAATATTGTCGCTGATTTTCAACAAGAGGAAATTAATCGGGCG 240
Db 17452 CGGCAGGAAGATCAAAATATTGTCGCTGATTTTCAACAAGAGGAAATTAATCGGGCG 17511
QY 241 AAAGTCAGCAGGAGGAAATTAATCTTTCGTAATTTAGCAAAACGGGCAATGCCAAC 300
Db 17512 AAAGTCAGCAGGAGGAAATTAATCTTTCGTAATTTAGCAAAACGGGCAATGCCAAC 17571
QY 301 GATGCCAGTATATCGCAAAACGCTTACGGTAATTAATGACAGTATTAATCCGAAAGTTCT 360
Db 17572 GATGCCAGTATATCGCAAAACGCTTACGGTAATTAATGACAGTATTAATCCGAAAGTTCT 17631
QY 361 GAAATTAAGGCATATATTAACCAAGTACGTAACGAGAAACAGACAGTTAGTGCAGAAA 420
```



```

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82757"
/translation="MATLSRLPIHVKMRGIGLTHALADISGLAFDRIPMTESDGT
FITAROPQWRPTSPPLHDLHLTPAGSSALVRPTDPTPODAPTEPWGHHPTARVA
PTAINQMLSGFSDVQLCWVGQPOLTRVYKHNAVPLGFAIGYFILINESLRDLOQ
KCPAGVQMEQFRPNLVSGVAVAMEDSKVRIGDVIIVVPCSRCLFTTVSPKQ
KHPSGEPALQAFRTAQDNGVDVQGLIIRNSGVIRVGDVEVIELATAPAKAYGATT
LDGVTPEKHDPDQSVTIDMOGQTFPCGNQVLEQLBQXGIRIPYSCRAGICGCCRIR
LLEGVSPLEKSGAMQDDITLSCSCVPTALRLEN"
/complement(4737..4955)
/gene="STY1081"
/notes="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
/gene="STY1082"
5924..8032
/gene="STY1082"
5924..8032
/notes="Similar to Haemophilus influenzae hypothetical
protein h10116 H10116 SW:YCBY HABIN (P44524; P43945) fasta
scores: E() : 0, 60.3% id in 710 aa
orthologue of E. coli YCBY ECOLI; Fasta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82756"
/db_xref="SPTRMBL:Q82756"
/translation="MNSLFPASTARGLELLIELEKLGAVGCVQVQGVHFGQDTRLI
YQSLMWSRLAIRILPMGECKVSDLDLYFVQAIINTEIPNPGATFVPSGLNDTI
RNSQYGAWKVDAIVDAFTAKRLPRPNVDRSPDLRIIVMVKESAIADLSDGLH
LRGYRDTGLAIPKETLAAATVMSRQPCRLIDPMGSGTLLIEAMWMTDAPRL
HKGHWGSGNAQDBAIIWQYKAEQRTARKGLBISHFPGSDARVIRAKRNS
RAGIGELTFEVKQVQALSNLPKPGYGTVISNPYSGRSDPALIALHLRLTKMK
NOFGWMLSLFSAAPDLGLQLRADQPKAKNPLDCVQKNYHIAETLADSKPATVA
EDYANRLKNKLEKEMAROGIECYRLYDADLEPVNAVRYGDMAVIOYAPKATV
DAKARORLFDIIATLSVLGIIPNKVLTKTRESQKGNQYOKSKSEPFLESEYNA
RLWNLTDYLDLTLGFLDRIIRARMLGEMSKKDLNLESYTSGASVYAGGAGANTT
VDMSRTYEMAEERNLKNGLSGRRARLQADCLMLKREANQPLIPIPTPESNSKR
MEESFDVQRDHVALMKDKRLRKGGTTFMSNNRGRFMDLEGLAEGLTAQETQKT
LSPDFARROIHNCMLIRAA"
5927..7060
/gene="STY1082"
/notes="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
/gene="STY1082"
6506..6541
/notes="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/notes="STY1083"
8045..9952
/notes="Fasta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Fasta hit to YJK_ECOLI (554 aa), 34% identity in 524 aa
overlap"
Query Match 98.9%; Score 451.2; DB 1; Length 254050;
Best Local Similarity 99.3%; Pred. No. 1,2e-123;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 61 GCGCAATATGATCTGGCTCGTTCAGATATATTTTGGCGTAATGAATTAAGCAG 120
DB 88469 GCGCAATATGATCTGGCTCGTTCAGATATATTTTGGCGTAATGAATTAAGCAG 88528
QY 121 TCTTCATTTATTCAGCGCGCCATTATTTGTCAGATCGGCACGATTAATAGTCCAGAGTA 180
DB 88529 TCTTCATTTATTCAGCGCGCCATTATTTGTCAGATCGGCACGATTAATAGTCCAGAGTA 88568
QY 181 CGCGAGGAGATCAAACTATTTGCTTATTTTCACAGAGGAGGAAATTAATCGGCG 240
DB 88589 CGCGAGGAGATCAAACTATTTGCTTATTTTCACAGAGGAGGAAATTAATCGGCG 88648
QY 241 AAATTCACGAGGAGGAAATTAATTCCTGCTTATTTAGCAAAACGGGAAATCGCAC 300
DB 88649 AAATTCACGAGGAGGAAATTAATTCCTGCTTATTTAGCAAAACGGGAAATCGCAC 88708
QY 301 GATCCAGTATATCGCAAGCGCTTACGTAATATGTCAGATTAATTCAGAAAGTTCT 360
DB 88709 GATCCAGTATATCGCAAGCGCTTACGTAATATGTCAGATTAATTCAGAAAGTTCT 88768
QY 361 GGAATTAAGCCAAATATTATCCAGTACGTACGAGAAACAGCAATTTGATGACAGAA 420
DB 88769 GGAATTAAGCCAAATATTATCCAGTACGTACGAGAAACAGCAATTTGATGACAGAA 88828
QY 421 CAGTCGATATGCTATTTGCGGTACCCAGCCTAA 456
DB 88829 CAGTCGATATGCTATTTGCGGTACCCAGCCTAA 88864

```

```

RESULT 5
LOCUS AE016840/c 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi Ty2
Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 (bases 1 to 301983)
Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.,
Comparative Genomics of Salmonella enterica Serovar Typhi Strains
Ty2 and CT18
J. Bacteriol. 185 (7), 2330-2337 (2003)
JOURNAL MEDLINE 22531367
PUBMED 12644504
2 (bases 1 to 301983)
Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.,
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1. 301983
Location/Qualifiers
Typhi Ty2
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
/complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
/complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
/notes="corresponds to STY1225 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
CDS

```

[illegible]

	Query Match	Best Local Similarity	76.5% 85.3%	Score 348.8; Pred. No. 1.4e-91;	DB 1;	Length 2889;
	Matches 389;	Conservative 0;	Mismatches 67;	Indels 0;	Gaps 0;	
Qy	1	ATGAAAAACAATTGTTATTATGATGTCACATACCTGGGTCGGCTGGAGTGCACC	60			
Db	1623	ATGAAAAACAATTGTTATTATGATGTTAACGTAAGTGGGTCGGCTGAAATTGCTTCC	1682			
Qy	61	GCACAATATTATGATCTGGCTCGTTCAGATATATATTTTGCGGTAAATGAATTAAAGCAAG	120			
Db	1683	GCACAAGTATTATGATTTAGCGCATTCGCAATATAACTTTGCCGTAAATGAATTAAAGTAAAG	1742			
Qy	121	TCTTCATTATATCAGCGCGGCATTTATGTGCACAGTCGCGACGATTAATGTGCCAGACTA	180			
Db	1743	TCTTCATTATATCAGCGCACGCATTTATGTGCACAGTGGCGCACAAATTAATGTGCCAAAGATG	1802			
Qy	181	CGCAGAGAGGATCAAAATCTATGTCCGTTATTTCCACAGAGAGAGCAAAATTAATCGGGCG	240			
Db	1803	CGCAGAGAGGAGCTCAAAATCTTTGTGCGTAAATTTCCACAGAGAGGTGAAGTAACTCGGGCC	1862			
Qy	241	AAAGTCGACGAGCGAGCAATATATCTTCCGATATGATGATGATGATGATGATGATGATGATG				

QY 361 GGAATTAAGGCCAATTATTACCGTAGTACGACAGAAAAACAGCAGTTGTAGTGCAGAAA 420
 Db 1993 GGAATTAAGCAATATATTACTAGTATNGTACCCAAAAACAGCAGTTGTAGTGCAGAGA 2042
 QY 421 CAGTCGATATGGGTATTCGCGTCCACCCAGCCTAA 456
 Db 2043 CAGTCGCAATGCGCAATGCGCGTTATTCAACGCTAA 2078
 RESULT 8
 AE015131
 LOCUS
 DEFINITION Shigella flexneri 2a str. 301 section 94 of 412 of the complete
 ACCESSION AE015131 AE005674
 VERSION AE015131.1 GI:24051313
 KEYWORDS
 SOURCE
 ORGANISM Shigella flexneri 2a str. 301
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Shigella.
 1 (bases 1 to 10370)
 REFERENCE
 AUTHORS Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H.,
 Liu, H., Yang, F., Yang, F., Qu, D., Zhang, X. B., Zhang, J. Y., Yang, G. W.,
 Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P.,
 Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L.,
 Qiang, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J.
 TITLE Genome sequence of Shigella flexneri 2a: insights into
 pathogenicity through comparison with genomes of Escherichia coli
 K12 and O157
 Nucleic Acids Res. 30 (70) 4432-4443 (2002)

FEATURES

TITLE Direct Submission of, Wei, J. M. and Hou, Y. D.
JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
 of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R.
 China
FEATURES Location/Qualifiers
 1..10370
 /organism="Shigella flexneri 2a str. 301"

gene /mol_type="genomic DNA"
/strain="301"
/rettype="2a"
/db_xref="taxon:198214"
complement(127..214)
/product="cRNA-Ser"
/note="anticodon: CGA"
410..1387
/gene="ycdW"
/locus_tag="SF1027"
/pseudo
410..1387
/gene="ycdW"
/locus_tag="SF1027"
/note="similar to Escherichia coli K12 putative
dehydrogenase gi: 1787270 (326 aa). BLAST with identity of
99% in 326 aa. This CDS contains an in-frame stop codon.
The sequence has been checked and is believed to be
correct."
/pseudo
/codon_start=1
/transl_table=11
1442..2179
/gene="ycdX"
/locus_tag="SF1028"
1442..2179
/gene="ycdX"
/locus_tag="SF1028"
/note="Residues 1 to 245 of 245 are 99 pct identical to
residues 1 to 245 of a 245 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287168.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42652.1"
/db_xref="GI:24051314"
/translation="MYVDLHMHTVASTHVAVSTSDYIAOKKGKIKLPAITDGPDM
EDAPHHMFRIMRIMPRVDGVLIGTEANIKNVDGSDSGKMPDSDLIIGFHE
FVPAPHDKATQTQAMISTIASGNVHITISHPNPKETIVKVAEAKAHQVALEINNS
FSLHSRKGSEDCRAVAAVADAGGVALSDSHAFWGEFECLKILDVDFPLER
ILNVSPRLNLFBSRGMAPIAEFADL"
2203..2757
/gene="ycdY"
/locus_tag="SF1029"
2203..2757
/gene="ycdY"
/locus_tag="SF1029"
/note="Residues 1 to 184 of 184 are 99 pct identical to
residues 1 to 184 of a 184 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287169.1"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase component"
/protein_id="AAN42653.1"
/db_xref="GI:24051315"
/translation="MNERSTICRVIGSGYYRQPDRLVPIFLIREKLANMPLEQ
DELLRLQKSCMTQVSNADYNALFTIGDECAVPYRSAAVEDATEEVAFLSEGMEL
ADTPADHIGTLILAAWLEDDSTEDSESLPTLSEYLLPWCAGFLGKVEAHATTPEW
RTMAPLTRDAISAMWDELDEDSSE"
2811..3350
/gene="ycdZ"
/locus_tag="SF1030"
2811..3350
/gene="ycdZ"
/locus_tag="SF1030"
/note="Residues 1 to 179 of 179 are 98 pct identical to
residues 1 to 179 of a 179 aa protein from Escherichia
coli K12 ref: NP_415554.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42654.1"
/db_xref="GI:24051316"

gene /translation="MCSISPMARFAAFMRGNILISTATTTGIIISGIMGVAVSIL
SNAGFLGTAFAFACPOGGLKLAISATLISGVYMANVITIGSLAPLHEITGYYTG
IVAFSLKTIQAKQLISFVPGTFICACATFAGQGWKVLPSLALGLVGYMKNSGLW
LAARSACTAHREQBIKNKA"
complement(3414..4189)
/gene="csqG"
/locus_tag="SF1031"
/pseudo
complement(3414..4189)
/gene="csqG"
/locus_tag="SF1031"
/note="similar to Escherichia coli K12 curlI production
assembly-transport component, 2nd curlI operon gi: 1787274
(278 aa). BLAST with identity of 97% in 257 aa. This CDS
contains an in-frame stop codon. The sequence has been
checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11
complement(4454..5662)
/locus_tag="SF1032"
complement(4454..5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to
residues 1 to 402 of a 402 aa protein from plasmid R100
ref: NP_052925.1"
/codon_start=1
/transl_table=11
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCSLDIHDLSLYOCPEHILKRNLSUTLACHALLDCTLTTEL
GRNLPTARTKNTIKRIDRLGNRHLKERVAVRMASFICSGNTMFTIVADSDIR
EQRKLWVLRASVALHGSAVTLYERAFPEQCSKKAHQFLADLASILPSNTPLIVS
DAGEKVPWYKSVLEKLGWMLSRVAGKYQVADLGAENKPDLSNLDMSSSHSKTGYKR
LTKSNPISCOLILYKRSKGRKQNRSTRHCHHSPTIVSAAKEPMIATNLVPEIR
TPKQLVNIVSKRMQIERTFPDLKSPAYGIGLRHSRTSSSEFPDMLIATMLQTCWL
AGVIAQOGMDKHQIANTVRRNVLSIVRLGMEVLRHSGYITTEDSDLVATLTATLQNL
FTHGIVLQKL"
complement(5995..6384)
/gene="csqB"
/locus_tag="SF1033"
complement(5995..6384)
/gene="csqB"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to
residues 1 to 129 of a 129 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287172.1"
/codon_start=1
/transl_table=11
/product="curlI production assembly-transport component,
2nd curlI operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRYLRWVAAEFLPAAGNLHAEVVEVPGILTDHTVSIGHPY
RABSDKESDYTGNTITNERPSAAWGSMTITTVQDVIFQFLPLPKRDSKTVVFL
IQTEBALNRQINQALSTDDLADDE"
complement(6389..7069)
/gene="csqD"
/locus_tag="SF1034"
complement(6389..7069)
/gene="csqD"
/locus_tag="SF1034"
/note="Residues 11 to 226 of 226 are 100 pct identical to
residues 1 to 216 of a 216 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=11
/product="putative 2-component transcriptional regulator
for 2nd curlI operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="MRSIKKMGFMFMFNEVSHIGHDTLLITKPSLQATVALLQHLKQSL

ATGKTHNIGRSLDSSGSIILLDMWADKLIHYMODTLRKNKNNIKILLNTPED
 VPDIDEMPHINGVYAMEDORVWVGCVGECYCTOKLAVLTHSNNIRYNS
 TESALTHREKEIKLIRIGASNNELARSIFISENTVKHLINLFRKLAIVKRIQAVS
 WANDNLRR"

gene
 7766..8248
 /gene="csbB"
 /locus_tag="SF1035"
 7766..8248
 /gene="csbB"
 /locus_tag="SF1035"
 /note="Residues 10 to 160 of 160 are 99 pct identical to
 coli O157:H7 EDL933 ref: NP_287175.1"
 /codon_start=1
 /transl_table=11
 /product="minor curliin subunit precursor, similar to CsgA"
 /protein_id="AA042658.1"
 /db_xref="GI:24051320"
 /translation="MYDOVQDNNKMLLFMLTILGAPGIAAAGYDLANSRYNFAV
 NEIKSSFEVQAAIIQAGTNNASQIRGGSGTLAVVORSSSRRAKIDGTQDYNLAVI
 DQAGSANDASISOGAVGNTAMIIQKSGSNKANTQYGTOKTAVVOROSQMAIRVYOR

CDS

Query Match 73.3%; Score 334.4; DB 1; Length 10370;
 Best Local Similarity 83.3%; Pred. No. 2.6e-87;
 Matches 380; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATGTTTATATATGATTTGACAATAGTGGTGGCGGATTTGCAAC 60
 DB 7793 ATGAAAAACAATGTTTATATATGATTTGACAATAGTGGTGGCGGATTTGCAAC 60
 QY 61 GCGACAAATTAATGATCTGGCTCGTTCAGATTAATTTTGGCGTAATGAATTAAGCAAG 120
 DB 7853 GCACAGGTTATGATTTAGCTAATTCAGATTAATTTTGGCGTAATGAATTAAGCAAG 120
 QY 121 TCTTCATTATACAGGGCGCATTTATTTGTCAGATGCGCAGCATATAGTCCAGAGTA 180
 DB 7913 TCTTCATTATACAGGGCGCATTTATTTGTCAGATGCGCAGCATATAGTCCAGAGTA 180
 QY 181 CGCCAGAGAGATCAAACTATGTCGTTATTTTCAAGAGAGAGAAATATTCGGCGC 240
 DB 7973 CGCCAGAGAGATCAAACTATGTCGTTATTTTCAAGAGAGAGAAATATTCGGCGC 240
 QY 241 AAATGCAAGCAGCAGGCAATTAATTAATTTGCGATTTATTTGAGCAAGCGGCAATGCCAAC 300
 DB 8033 AAATGCAAGCAGCAGGCAATTAATTAATTTGCGATTTATTTGAGCAAGCGGCAATGCCAAC 300
 QY 301 GATCCAGATTAATTCGCAAGCGCTACCGTATAGTGCAGTATTTATCCGAAGGTTCT 360
 DB 8093 GATCCAGATTAATTCGCAAGCGCTACCGTATAGTGCAGTATTTATCCGAAGGTTCT 360
 QY 361 GGAATTAAGGCAATTAATTAATTCGATGCGTACGAGAGAGAGAGAGAGAGAGAGAG 420
 DB 8153 GGAATTAAGGCAATTAATTAATTCGATGCGTACGATGCGTACGAGAGAGAGAGAGAGAGAG 420
 QY 421 CAGTATTAAGCAATTAATTAATTCAGATTAATTTGAGCAAGCGGCAATGCCAAC 8212
 DB 8213 CAGTATTAAGCAATTAATTAATTCAGATTAATTTGAGCAAGCGGCAATGCCAAC 8212

RESULT 9
 ECCSGABDG 4680 bp DNA linear BCT 07-Jul-2002
 LOCUS E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
 DEFINITION X00754.1 GI:1147558
 ACCESSION X00754.1 GI:1147558
 VERSION csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
 ORF C gene.
 KEYWORDS Escherichia coli
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCES
 1 Hammar, M., Arngvist, A., Bian, Z., Olsen, A. and Normark, S.

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 8817489
 2 (bases 1 to 4680)
 AUTHORS
 TITLE
 JOURNAL
 Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet
 Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
 SWEDEN

FEATURES
 source
 Location/Qualifiers
 1..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:562"
 /map="23.15 minutes"
 1..3140
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="MC4100"
 /db_xref="taxon:562"
 3141..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="W3110"
 /db_xref="taxon:562"
 <1..>4680
 /insertion_seq="IS2 (partial)"
 complement(133..966)
 /gene="csgG"
 complement(133..966)
 /gene="csgG"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 production"
 /protein_id="CA62277.1"
 /db_xref="GI:1147559"
 /db_xref="GOA:P52103"
 /db_xref="SWISS-PROT:P52103"
 /translation="MORFLVAVWLISGCLTAPPKAAPTLMRAQSYKDLTHPA
 FPKTRFVSVYNQDETFQKPYAPSNFSTAVPQATAMVYALKDSKFTLEKQGLQ
 IAGDYQYQDQIAVNLRVNVSSTGELISSVSTSTTISYVQGVFRFIDYORLLGE
 VGTENEPVMLCLMSAIEITGVIFILINDIRGLMDLQNKARQNDILVKKYHMSVPE
 S"

source

repeat_region

gene

CDS

gene

CDS

gene

CDS

Expression of two csg operons is required for production of
 fibronectin- and congo red-binding curli polymers in Escherichia
 coli K-12
 Mol. Microbiol. 18 (4), 661-670 (1995)
 5641468
 8817489
 2 (bases 1 to 4680)
 AUTHORS
 TITLE
 JOURNAL
 Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet
 Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
 SWEDEN

Query Match	73.0%;	Score 332.8;	DB 1;	Length 4680;
Best Local Similarity	83.1%;	Pred. No. 7.5e-87;		
Matches 379; Conservative	0;	Mismatches 77;	Indels 0;	Gaps 0;

RESULT 10	
LOCUS	AE000205
DEFINITION	AE000205 10346 bp DNA linear BCT 01-DEC-2000
ACCESSION	AE000205
VERSION	1
KEYWORDS	genome.
SOURCE	AE000205.1 GI:1787265
ORGANISM	Escherichia coli K12
REFERENCE	Escherichia coli K12
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
	1 (bases 1 to 10346)
	Blattner,F.R., Plunkett,G. II, Bloch,C.A., Berna,N.T., Burland,V.,
	Riley,M., Collado-vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
	Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.Y.,
	Mau,B. and Shao,Y.
	The complete genome sequence of Escherichia coli K-12
	Science 277 (5311), 1453-1474 (1997)
TITLE	97426617
JOURNAL	9278503
REFERENCE	2 (bases 1 to 10346)
AUTHORS	Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUN-1997) Guy Plunkett III, Laboratory of Genetics,
	University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
	Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
	608-263-7459
	3 (bases 1 to 10346)
	Blattner,F.R.
	Direct Submission
	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
	University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
	Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
	608-263-7459
	4 (bases 1 to 10346)
	Plunkett,G. III.
	Direct Submission
	Submitted (13-OCT-1998) Laboratory of Genetics, University of

COMMENT

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: markborov@gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Belyyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

```

source          location/Qualifiers
                1..10346
                /organism="Escherichia coli K12"
                /mol_type="Genomic DNA"
                /strain="K12"
                /sub_strain="MG1655"
                /db_xref="taxon:8333"
                /note="153"
                /complement(<1..483)

repeat_region  <1..5
                /note="factor Sigma70; predicted +1 start at 1094689"
                5..32
                /note="factor Sigma70; predicted +1 start at 1094716"
                69..392
                /gene="b1028"
                69..392
                /gene="b1028"
                /function="orf; Unknown"
                /note="0107; This 107 aa ORF is 27 pct identical (0 gaps)
                SW: P43098"
                /codon_start=1
                /transl_table=1
                /product="orf; hypothetical protein"
                /protein_id="AAC74113.1"
                /db_xref="GI:1787266"
                /translation="MEDGVQMRDYLAGDIDSPHEOVMTAKSEVAPSIKENTIS
                LHLHTYKMTESLDTPFIMPTVPVAFLEINSKPYLGMLNDPAGVDQKRWMCRAFSID
                AKVFR"
                275..299
                /gene="b1028"
                /note="factor Sigma32; predicted +1 start at 1094983"
                389..1375
                /gene="ycdU"
                /note="synonym: b1029"
                389..1375
                /gene="ycdU"
                /function="orf; Unknown"
                /note="0328; This 328 aa ORF is 19 pct identical (4 gaps)
                SW: P47562"
                /codon_start=1
                /transl_table=1
                /product="orf; hypothetical protein"
                /protein_id="AAC74114.1"

CDS
                /db_xref="GI:1787267"
                /translation="MIPDYLFPIRPDKRLIYVAILGLIGFVKNGAFPPSPEDI
                GVSSGLILVYNFIFDLKAYMAKCTKIDSSWRKQNMRIELFTQPVAFSL
                LIMSAMSWMGLYQILPSLYALPLISLIGPLVILFLRMRTSYVKQVALSVAKRYK
                SLTRVLLSVICISVNVLLTISPLRNSDSVTEGQWLTFKSIILALICGVLAINTLF
                FLRFSKRYAFIGRFLQRIIDTFSSNSALSTFPKALMLRLILVIEVMITLVSVL
                ALVEMRIEMPAVTLCTVPCLLIYFFYCRFLMNDPMMACMVFWRWHFHK"
                1253..1280
                /gene="ycdU"
                /note="central position to predicted promoter: -104"
                /bound_moiety="Deor predicted site"
                1253..1280
                /gene="ycdU"
                /note="No predicted promoter"
                /bound_moiety="Deor predicted site"
                1337..1365
                /gene="ycdU"
                /note="factor Sigma70; predicted +1 start at 1096049"
                1494..1745
                /gene="b1030"
                1494..1745
                /gene="b1030"
                /function="orf; Unknown"
                /note="083; This 83 aa ORF is 25 pct identical (7 gaps)
                to 177 residues of an approx. 616 aa protein ALBU_PABIT SW:
                P49065"
                /codon_start=1
                /transl_table=1
                /product="orf; hypothetical protein"
                /protein_id="AAC74115.1"
                /db_xref="GI:1787268"
                /translation="MGGCTRCVQYTKKKEVLYEELFFKYGEGIDSLRSPCGQPAHC
                VHGLSNMLSPVUPRSGLLIPVCAIYHKKARTFVALL"
                1513..1926
                /gene="ycdV"
                /note="synonym: b1031"
                1513..1926
                /gene="ycdV"
                /function="putative structure; Not classified"
                /note="0137; This 137 aa ORF is 27 pct identical (1 gap)
                to 47 residues of an approx. 80 aa protein RS37_YEAST SW:
                P05759"
                /codon_start=1
                /transl_table=1
                /product="putative ribosomal protein"
                /protein_id="AAC74116.1"
                /db_xref="GI:1787269"
                /translation="MCNIRKSPYFRTSSILNAAVRGGLTRFAPRAGSPDLTAFTVCP
                TGRQSTPGRGSSPRCVQYTKKAPVLSYELFFRYGEGIDSLRSPCGQPAHCVHGL
                SNWLSPVVPRSGLLIPVCAIYHKKARTFVALL"
                complement(2111..2198)
                /gene="bexX"
                /note="synonym: b1032"
                complement(2111..2198)
                /gene="bexX"
                /product="tRNA-Ser"
                /note="anticodon: GGA; CG Site No. 17878"
                2269..2296
                /function="RNA; tRNA"
                2269..2296
                /note="factor Sigma70; predicted +1 start at 1096980"
                2338..2365
                /note="factor Sigma70; predicted +1 start at 1097049"
                2355..2383
                /note="factor Sigma70; predicted +1 start at 1097067"
                2393..3370
                /gene="ycdW"
                /note="synonym: b1033"
                2393..3370
                /gene="ycdW"
                /function="putative enzyme; Not classified"
                /note="0325; This 325 aa ORF is 32 pct identical (2 gaps)
                to 177 residues of an approx. 312 aa protein YPRA_CORGL
                SW: P45637"

```

```

/codon_start=1
/strand1_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="MRINFAQRFSGVSDIIIFYHPTEDTQWMIETAKRKAIPQARVRAM
KSGNDSADYALVWHPVEMALGRDLKAVPLAGVDSILSKLOHPMLPSVLPFR
LEDGMEQMEVAVSOVLHWPFRREDYRIVQNSHMOPLPEYHREDTIGILGAVL
GSKVAOSITQWRFPRRCMSRTRKSPQSGAGRELSAFISQCRVLINLPLPRTY
GIINQDLKLPDGAITLNLKGVHVDLALADSGKRVGMLDVNRPPLPPESP
LMQHPRTVITTHVAITRPAVEYISRTIAQLKGRKVCQVDVRRGY"
3278..3308
/gene="ycdW"
/note="factor Sigma70; predicted +1 start at 1097992"
3383..3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396..3423
promoter

Query Match      73.0%; Score 332.8; DB 1; Length 10346;
Best Local Similarity 83.1%; Pred. No. 7,7e-87;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACGGTGGCGCTGGATTGCAAC 60
Db 8497 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACGGTGGCGCTGGATTGCAAC 8556
Qy 61 GCGACAATTATGATCTGCTCGTTCAAAATTTATTTGGCGTAAATGAATTGAACAG 120
Db 8557 GCACACAGTTATGATTTAGTTATTCAGAAATTAATCTTGGCGTAAATGAATTGAAG 8616
Qy 121 TCTTCATTATTAACAGCGCGCATTTTGTGTCAGTCGCGACGAGTAATAGTCCAGACTA 180
Db 8617 TCTTCATTATTAACAGCGCGCATTTTGTGTCAGTCGCGACGAGTAATAGTCCAGACTA 8676
Qy 181 CGCAGAGAGATCAAACTATTTGCTGCTTATTTCAAGAGAGAGAAATTAATCGGCG 240
Db 8677 CGCAGAGAGATCAAACTATTTGCTGCTGTTGCGCAAGAGAGATTAAGACACCGGCA 8736
Qy 241 AAATCCAGCCAGCGGCAATTTATTAATCTTGGCTATATGACAAACGGGCAATGCCAAC 300
Db 8737 AAATCCAGCCAGCGGCAATTTATTAATCTTGGCTATATGACAAACGGGCAATGCCAAC 8796
Qy 301 GATGCCAGTATATCGCAAGCGCTTACCGTATATAGTCAGCTATATTCAGAAAGTTCT 360
Db 8797 GATGCCAGTATATCGCAAGCGCTTATAGTATTAATCGGATGATTTATCAAAAGTTCT 8856
Qy 361 GGAATAAGGCAATATTACCGAGTACGTAACGAGAAACAGCAGTTGATGTCAGAA 420
Db 8857 GGTATATAAGCAATATTACAGATATGTACTCAAAAACCGCAATTGTATGTCAGAA 8916
Qy 421 CAGTCCGATATGCTATTCGCTGCTCAACCAAGCTTA 456
Db 8917 CAGTCCGATATGCTATTCGCTGCTGACACCAAGCTTA 8952

```

```

RESULT 11
LOCUS      D90741      15047 bp      DNA      linear      BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION  D90741 AB001340
VERSION     D90741.1 GI:1561509
KEYWORDS    Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
SOURCE      Escherichia coli K12
ORGANISM    Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
AUTHORS     Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A.,
            Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
            Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
            Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
            Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K.,
            Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T.

```

```

TITLE      A 718-kb DNA sequence of the Escherichia coli K-12 genome
JOURNAL    DNA Res. 3 (3), 137-155 (1996)
MEDLINE    97061202
PUBMED     8905232
REFERENCE   2
AUTHORS     Aiba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A., Horiuchi, T.,
            Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
            Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
            Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
            Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
            Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
            Yamamoto, Y. and Yano, M.
TITLE      The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL    Unpublished
REFERENCE   3 (bases 1 to 15047)
AUTHORS     Mori, H.
TITLE      Direct Submission
JOURNAL    Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info., 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail:hmori@gc.t.aisr-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project: The Japan E.coli genome DNA sequencing project
            Group: The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
            Aiba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A.,
            Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
            Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
            Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
            Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
            Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,
            Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
            Yamamoto, Y. and Yano, M.
            Headed by:
            Name: Takaishi Horiuchi
            Address: National Institute of Basic Biology, Okazaki, 444, Japan
            E-mail: kihori@nibb.ac.jp
            Information operator:
            Name: Hirotsada Mori
            Address: NARA Institute of Science and Technology,
            Ikoma, 630-01, Japan
            E-mail: hmori@gc.t.aisr-nara.ac.jp
            URL:
            The Japan E. coli genome database
            http://bsw3.aisr-nara.ac.jp.
            Location/Qualifiers
                1..15047
                /organism="Escherichia coli K12"
                /mol_type="genomic DNA"
                /strain="K12"
                /db_xref="taxon:83333"
                /map="23.7-24.0 min"
                /clone="Kohara clone #231"
                /note="Nucleotide position 1097101-1112147 from the
                initiation site of Thra (0 min.). ~This clone is from
                Kohara lambda miniset library."
                <1..276
                /note="unamed protein product; ORF_ID:0230#5"
                /codon_start=1
                /transl_table=11
                /protein_id="BAJ5822.1"
                /db_xref="GI:4062605"
                /translation="DIPFSEENALSTFPKAPLWRLPILIVIEWMITLVSUATLVE
                WRIFEAFLCTVPCILTYFFYCRFLMHNDFMMACDMFRMGHFNK"
                1152..2090
                /note="ORF_ID:0231#1
                similar to SwissProt Accession Number P45637"
                /codon_start=1

```

```

gene      Complement(4977, .5393)
          /gene="ccsgf"
          complement(4977, .5393)
CDS       /gene="ccsgf"
          /note="ORF_ID:0231#6
          similar to SwissProt Accession Number P52104"
          /codon_start=1
          /transl_table=11
          /product="Curl1 production assembly/transport component
          Csgf precursor."
          /protein_id="BAA35828.1"

```

aligned to SwissProt Accession Number P52106"
 /codon_start=1
 /transl_table=11
 /product="Probable csgAB operon transcriptional regulatory
 protein. "
 /protein_id="BAJ35830.1"
 /db_xref="GI:4062613"
 /translation="MFENEVHSIHGHTLLITKSLQATALLQHLKQSALATGKHTNIQ
 RSLDDISDSGSIILLDMEMDKLIHWMDITLSPKNNITKILNLPEDYPRDIENNE
 HINGVYSEMDERVNGLOSVLRGECTFQGLASVTLISGNRVNFTSALLTHPE
 KELLNKKRLGNSNNEIARSLPISENTVKTLYNLFKCIQAVKQRTQAVAMNDNLR"
 7217. 7672
 /gene="csgB"
 /gene="csgB"
 7217. 7672

Query Match	73.0%	Score 332.8;	DB 1;	Length 15047;
Best Local Similarity	83.1%;	Ped. No. 7.8e-87;		
Matches 379;	Conservative	0;	Mismatches 77;	Indels 0;
				Gaps 0

Db	7217	ATGAAAAA	CAAAATGTTATTTATATGATGTGA	CAAAATCGGGTGCGCTCGGGATTGGCAAC	60
OY	61	GGACAAAT	TATGATCTGGCTCGTTACAAATTTAA	TTTTGGGTAAATGAATTAAGCAAG	120
Db	7277	GCAGCAGG	TATGATTTAGTACTATTCAAAATTTAA	CTTCGGGTAAATTAATGATTAAG	7336
OY	121	TCTTATTT	AATCAGCGCGCATTAATGGTCAGTCGGC	ACGATTAATAGTCGACAGTA	180
Db	7337	TCTTATTT	ATTAATCAGGCGACCATATTGGTCAAC	CTGGACTAATTAATAGTCTCAAGTA	7396
OY	181	CGCCAGAA	GATGTAAAACTATTTCCGTTATTTCA	CAAGAAAGGAGGAAATTAATCGGCG	240
Db	7397	CGGACGGG	AGGCTCAAACTTTTGCGGTTTGCCG	CAAGAAAGTATGATGACACACCGGCA	7456
OY	241	AAAGTCAC	CAGGAGGAAATTAATCTTGGCTATT	TAGGAAACGGGCATTCGCAC	300
Db	7457	AAGATTAC	CAGACAGAGATTTAACTTGCAATTTG	ATTTAGTCAGCGGGCACTGCGAAC	7516
OY	301	GATCCAGT	ATTATCGCAACCGCTTAACGTAATAG	TGCACTTTATTCAGAAAGGTTCT	360
Db	7517	GATCCAGT	ATTTCGAAAGGTGTTATGTGTAATAC	TGCGATGTTATTCAGAAAGGTTCT	7576
OY	361	GGAATTA	AGGCAATATTACCCAGTACGGTAAG	CAAAAACAGCACTGTAGTCAGAAA	420
Db	7577	GGTAAATA	AGCAAAATTTTACAGATTTGTACTCA	AAAAACGSCAAATGTATGTGACAG	7636
OY	421	CAGTCCAT	TATGCTATTGCGGTCAACCAACGCT	TAA	456

DB 7637 CAGTGCMAATGGTATTGCGTGCAGACACGTTAA 7672

RESULT 12
LOCUS AE016981 292504 bp DNA linear BCT 22-APR-2003
DEFINITION *Shigella flexneri* 2a str. 2457T section 4 of 16 of the complete genome.
ACCESSION AE016981 AE014073
VERSION AE016981.1 GI:30040616
KEYWORDS
SOURCE *Shigella flexneri* 2a str. 2457T
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Shigella*.
REFERENCE 1 (bases 1 to 292504)
AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
TITLE Complete Genome Sequence and Comparative Genomics of *Shigella flexneri* Serotype 2a Strain 2457T
JOURNAL Infect. Immun. 71 (5), 2775-2786 (2003)
PUBMED 12704152
REFERENCES 2 (bases 1 to 292504)
AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
SOURCE location/Qualifiers
1..292504
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
complement(223..441)
/gene="infA"
/locus_tag="S0884"
complement(223..441)
/gene="infA"
/locus_tag="S0884"
/function="factor; Proteins - translation and modification"
/note="residues 1 to 72 of 72 are 100.00 pct identical to residues 1 to 72 of 72 from *Escherichia coli* K-12 : B0884"
/codon_start=1
/transl_table=1
/product="protein chain initiation factor IF-1"
/protein_id="AAP16348.1"
/db_xref="GI:30040617"
translation="MAKEDNIEMOQIVLETLPTMTFRVLENGHVVAHSKGRKKNYIRITGDKVTELTPEYLSKGRIVFRSR"
complement(726..1430)
/gene="atf"
/locus_tag="S0885"
complement(726..1430)
/gene="atf"
/locus_tag="S0885"
/function="enzyme; Aminoacyl tRNA synthetases, tRNA modification"
/note="residues 1 to 234 of 234 are 100.00 pct identical to residues 1 to 234 of 234 from *Escherichia coli* K-12 : B0885"
/codon_start=1
/transl_table=1
/product="tycyl, phenylalanyl-tRNA-protein transferase"
/protein_id="AAP16349.1"
/db_xref="GI:30040618"
translation="WRLVQLSRHSIAFSPREGALREPNGLLALGGLSPARLLAYOR

GIPIPFSGDPIILMSPDPAVLVMPESLIHSRSMKRPHKSPVRYTMNVAFGVIEGC
ASDRSGTWTITRGVVEAYRHLELGHASHLEWREDELVGCMVYVAGTCTFCGSMPS
RMENASTALLVFCBEFTIGHGKLIIDQVLNDHTASJACACIPRDIYLTNYLNRLEGR
LPNNFWPRLCFSPQE"
complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
/function="transport; Transport of small molecules; Other"
/note="residues 1 to 573 of 573 are 97.03 pct identical to residues 1 to 573 of 573 from *Escherichia coli* K-12 : B0886"
/codon_start=1
/transl_table=1
/product="ATP-binding component of cytochrome-related transport"
/protein_id="AAP16350.1"
/db_xref="GI:30040619"
translation="WRALPPLALYKHKMMLSGIVLAVTLASIGLLTSGMPLS
ASAVAGVAGLYSENYMLPAAGVRCALITRAGRFEELVSDATFRLQHLRIYFSK
LLPLSPAGLARVQSGELNVDVDTLDHLYLIVISPLVGAFAVIMVITIGLSLDF
TLAFTLSVMLLTLPLMPLEFRAGKSTGNLTILRGYQQLTAMOGALETIFGA
SDRYRQLENTETIOMLEAORQSELTALSOIMLIGALAVIILMMASGGVGNAP
GALIALVFCALAAFEALAVYTAFOHLYGVYASAVASIDLTOKREVPPTDTRVA
DRSLTLDVQFTPEOSQDLKITSIQVNAHEBITALRGCKSTLQQLTAMPD
QOGELILNDSPILASLAAURQITSVPRHLEFSAULRNLLASRGSSEALSL
RRLVLETLIEDAGLNSWLGSGRQLSGEELRLIALTLTLDADVLVLDTEBEDLDT
TESQILLELLAKMREKTVLWYTHRLRGLSRFQQLIIVNDQGIIRGTHAELLARQRY
YQPKQGI"
complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"
complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"
/function="transport; Not classified"
/note="residues 1 to 588 of 588 are 92.85 pct identical to residues 1 to 588 of 588 from *Escherichia coli* K-12 : B0887"
/codon_start=1
/transl_table=1
/product="ATP-binding component of cytochrome-related transport"
/protein_id="AAP16351.1"
/db_xref="GI:30040620"
translation="MNSKROKELTRMLKQOSVISQWMLNLSRLGPFVGLIIIAQWF
MARLQIMIMENIPREALLPFTLVITFLRAVWMLREBVGTHAGOHIRFAIRQV
LDRLOQGPAMIOGKPGSWATYLBQIDDMHVVYAYLQMLAVSVPILYIATRP
SNWALILGTAFLPLPFWALVGMGADANRRNRLYALRSGHFLDRRMEYLRIF
GREGARIESIRASSEDPRQRTMEVLRALFSSGILFEFTSLISALVAVYGFSTGL
DPGHYDTGVTLAAGFLALILAPERFQPLRDLGTFYHAKAGAVGADSLKTFEMPLH
PQREAEALATLDTPIEADLFTISPEGKTLAGLNTLTPAGCAVAVGSGSSSL
LNLISGLSYOGSLRIVGIELRDISPSMKRHSWQGNLPAAKLRDNLVLRPAA
SEQELQAALNNAVSETPLRPGQVDPYVQDQARLSVCGAQRARAVARALLPESLL
LDEPASLDHSHSEORVMEALNARSROTTLVTHQLDLADMDVITWQDSRIIEQGR
YAEISVAGSPATLLAHROEI"
complement(5083..6048)
/locus_tag="S0888"
/gene="trxB"
/locus_tag="S0888"
complement(5083..6048)
/gene="trxB"
/locus_tag="S0888"
/function="enzyme; 2-Deoxyribonucleotide metabolism"
/note="residues 1 to 321 of 321 are 95.95 pct identical to residues 1 to 321 of 321 from *Escherichia coli* K-12 : B0888"
/codon_start=1
/transl_table=1
/product="thioredoxin reductase"
/protein_id="AAP16352.1"
/db_xref="GI:30040621"

QY	1	ATAAAAACAATTGTTATTATTAGATTGACATACTGGGTGGCCCTGGGATTGGCAAC	60
Db	44773	ATAAAAACAATTGTTATTATTAGATTGACATACTGGGTGGCCCTGGGATTGGCAAC	4483
QY	61	GCACAAATATGATCTGGCTGGTCAGAAATTAATTTTGGCGTAATGAATTAAGCAAG	120
Db	44833	GCAGCAGGTATGATTACTTAATTCAGAAATTAATTTTGGCGTAATGAATTAAGCAAG	44899
QY	121	TCTTCAATTAATCAGCGGCGCAATTAATGGTCAAGTCGCGACGAAATTAAGCCAGAGTA	180
Db	44893	TCTTCAATTAATCAGCGACCGCAATTAATGGTCAAGTCGCGAATAATTAATGCTCAAGTA	44955
QY	181	CGCGAAGAGATCAAAAATATTTGTCGTTATTTTCACAAGAGAGAAATTAATCGGCG	240
Db	44953	CGCGAGGAGGCTCAAAAATTTTGACGGTGTGTCGAAAGAGGTATATCAACCGGGCA	45012
QY	241	AAAGTCGACCGAGCGGGAATTAATCTTTGCGTATATTGACCAACCGGCGAATGCGAAC	300
Db	45013	AAAGTTGATCGACGAGGAGATTATTAACCTTGTCATATATTGATCGAGCGGCAATGCGAAC	45072
QY	301	GATGCGAGTATATCGCAAGCGCTTAACGGTAATGTGCGACCTATTAATCCAGAAAGTTCT	360
Db	45073	GATGCGAGTATTTTGGCAAGGTCCTTAATGCTTAATGCGATGATTAATCCAGAAAGTTCT	45132
QY	361	GGAAATTAAGCGCAATTAATCCCAAGTACGTACGACAGAAAAACAGCAATTGTAGTCAGAAA	420
Db	45133	GGTAATTAAGCAAAATTAATTAACAGTATGTATCTCAAAAAACGCAATTTATGTCAGAGA	45192
QY	421	CAGTCGATATGGCTATTCGCGTCAACCAACGGTTAA	456
Db	45193	CAGTCGCAATGGCTATTGCGGTGACCAACCGTTAA	45228

```

/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
757. .1212
/gene="csgB"
757. .1212
CDS

```

Query Match	72.6%;	Score 331.2;	DB 1;	Length 1711;
Best Local Similarity	82.9%;	Pred. No. 2.2e+86;		
Matches 378;	Conservative	0;	Mismatches 78;	Indels 0.

QY	1	ATGAAAAACAATTCGTATTATTAATGATGTTGACAAATACTGGGTGCCTCGGAGTTGCACAC	60
Db	757	ATGAAAAACAATTCGTATTATTAATGATGTTAACAATACTGGGTGCCTCGGAGTTGCACAC	816
QY	61	GCACAAATTATGATCGTGGCTCGTCACAAATTAATTTTGGCGTAAATGAATTAAGCAAG	120
Db	817	GCAGCGGTATGATTTTAGCTTAATTCAGAAATTAATCTTGGCGTAAATGAATTAAGTAAAG	876
QY	121	TCCTTATTTAATCAGCGCGCCATTAATGTCACATCGGCACGATTAATGAGCCAGAGTA	180
Db	877	TCCTTATTTAATCAGCGCGCCATTAATGTCACATCGGCACGATTAATGAGCTCAAGTA	936
QY	181	CGCCAGAAAGATCAAAAACATAATGTCGGTATTTTCAACAAGAGAGAAATTAATCGGCGC	240
Db	937	CGCAGAGAGGCTCAAAACTTTTGGCGGTGTCGCGAAGAAAGTATGATCAACCCGGCA	996
QY	241	AAAGTCACACGAGCGAAATTAATATTGGCGATTTAGCAAAACGGGCATATGCCAAC	300
Db	997	AAGATTAACACGACAGAGATTTAATACCTTGCAATATTAATCAGCGGGCACTGCAAT	1056
QY	301	GATCCGATTAATCGCAACGCGCTTAACGTTAATGTCAGCTAATATTCAGAAAGCTTCT	360
Db	1057	GATCCGATTAATCGCAAGGTGCTTAATGTAATACTGCATGATTAATTCAGAAAGCTTCT	1116
QY	361	GGAATTAAGCCCAATATTACCCAGTACGGTAAGCCAGAAAAACAGCACTTATGTCAGAAA	420
Db	1117	GGTATTAACCAATATTACAGATATGGTATCTCAAAAAACGGCAATGTATGTCAGAGAA	1176
QY	421	CAGTCGATATGGCTAATTCGCGTCAACCCAACGCTAA	456
Db	1177	CAGTCGCAATGGCTAATTCGCGTCAACCAAGCTAA	1212

RESULT 15	AE005315				
LOCUS	AE005315	10190 bp	DNA	linear	ECT 21-MAR-2001
DEFINITION	Escherichia coli O157:H7 EDL933 genome, config 1 of 3, section 139 of 155.				
ACCESSION	AE005315	AE005174			
VERSION	AE005315.1	GI:12514572			

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds

(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-2

Perfect score: 456
Sequence: 1 atgaataacaatcgtatc.....ctgcgcaccacgaactaa 456Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002s:*
- 7: Geneseq2003as:*
- 8: Geneseq2003bs:*
- 9: Geneseq2003cs:*
- 10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	3 AAC64618	AAC64618 Salmonella
2	332.8	73.0	456	3 AAC64620	AAC64620 Escherichia
3	74.4	16.3	100	7 ACD68810	ACD68810 E. coli K
4	66.4	14.6	100	7 ACD68812	ACD68812 E. coli K
5	64.8	14.2	100	7 ACD68811	ACD68811 E. coli K
6	37.6	8.2	456	3 AAC64623	AAC64623 Agfa::PT3
7	37.6	8.2	2000	7 ADA71938	ADA71938 Rice gene
8	35.8	7.9	39	3 AAC64608	AAC64608 Agfa (SEF)
9	35	7.7	738	6 AAD32400	AAD32400 Chlamydia
10	35	7.7	105325	6 ABR94407	ABR94407 DNA encod
11	34.4	7.5	2006	9 ADC10081	ADC10081 Human NOV
12	34.4	7.5	18540	8 ACD26496	ACD26496 Human tum
13	34.2	7.5	456	3 AAC64629	AAC64629 Agfa::PT3
14	34	7.5	361	2 AAQ73066	AAQ73066 Agfa sequ
15	34	7.5	361	2 AAT74141	AAT74141 Salmonella
16	34	7.5	456	2 AAQ87467	AAQ87467 Agfa sequ
17	34	7.5	456	2 AAT74142	AAT74142 Salmonella
18	34	7.5	456	3 AAC64628	AAC64628 Agfa::PT3
19	34	7.5	456	3 AAC64622	AAC64622 Agfa::PT3
20	34	7.5	456	3 AAC64617	AAC64617 Agfa::PT3
21	34	7.5	456	3 AAC64625	AAC64625 Agfa::PT3
22	34	7.5	456	3 AAC64624	AAC64624 Agfa::PT3
23	34	7.5	2000	7 ADA71938	ADA71938 Rice gene

C	24	33.8	7.4	203654	7 ABX16034	ABX16034 Human gen
C	25	33.4	7.3	53585	2 AAX20251	AAX20251 Borrelia
	26	33.2	7.3	535	9 ADD29528	ADD29528 Mouse cae
	27	33	7.2	33	3 AAC64614	AAC64614 S. enteri
	28	33	7.2	3699	6 ABN70306	ABN70306 Streptococ
	29	33	7.2	3699	6 ABN70834	ABN70834 Streptococ
	30	33	7.2	3717	6 ABN66453	ABN66453 Streptococ
	31	33	7.2	110000	6 ABN71527_03	Continuation (4 of
	32	32.8	7.2	418	3 AAH30547	AAH30547 Human col
	33	32.8	7.2	1083	9 ADC10077	ADC10077 Human NOV
	34	32.8	7.2	1098	9 ADC10079	ADC10079 Human NOV
	35	32.8	7.2	1105	9 ADC10075	ADC10075 Human NOV
	36	32.8	7.2	1168	9 ADC10069	ADC10069 Human NOV
	37	32.8	7.2	1171	9 ADC10071	ADC10071 Human NOV
	38	32.8	7.2	1299	9 ADC10073	ADC10073 Human NOV
	39	32.8	7.2	1938	9 ADC10067	ADC10067 Human NOV
	40	32.8	7.2	2649	7 ADA53932	ADA53932 Human cod
	41	32.8	7.2	2843	7 ACC44405	ACC44405 Gene enco
	42	32.4	7.1	779	1 AAN40272	AAN40272 Sequence
	43	32.4	7.1	1914	1 AAE51585	AAE51585 Staphyloc
	44	32.4	7.1	1929	4 AAS54408	AAS54408 Staphyloc
	45	32.4	7.1	10732	3 AAA10594	AAA10594 Gene enco

ALIGNMENTS

RESULT 1
ID AAC64618 standard; DNA; 456 BP.
XX
AC AAC64618;
XX

26-FEB-2001 (first entry)

Salmonella enteritidis Agfa DNA sequence SEQ ID NO:2.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
vaccine; immune response; immunogen; ds.

Salmonella enteritidis.

MO200060102-42.

12-OCT-2000.

05-APR-2000; 2000MO-CA000356.

05-APR-1999; 99US-0127888P.

(UYVT-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay WW;

WPI: 2000-672631/65.

P-PSDB; AAB36342.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
which encodes foreign epitope or antigen, expresses recombinant Agfa
protein useful for eliciting immune response in animal.

Disclosure; Page 134; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
segment of the gene has been replaced by a segment of a foreign DNA
sequence which encodes a foreign epitope or antigen. Also described are:
(1) use of thin aggregative fimbriae (SBEF1/TAF) nucleation depended
assembly system of strains of Salmonella, Escherichia coli and
Enterobacteriaceae for the production of fimbriae comprising recombinant
Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
directing recombination of a recombinant gene into the chromosome of the
homologous species; (3) directing recombination of a recombinant gene
back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 8,6e-135; Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAAAACAATTTGTTATTTATGATGTTGACAACTAGTGGTGGCTGGGATTGCAAC 60
 1 ATGAAAAACAATTTGTTATTTATGATGTTGACAACTAGTGGTGGCTGGGATTGCAAC 60
 61 GCGCAAAATTTGATCTGGCTGCTTGAATATTTTGGTAAATGAATTAAGCAAG 120
 61 GCGCAAAATTTGATCTGGCTGCTTGAATATTTTGGTAAATGAATTAAGCAAG 120
 121 TCTTCATTTATTCAGGGGGCCATTATTTGGTCAAGTGGGACGATTAATAGTCCAGATA 180
 121 TCTTCATTTATTCAGGGGGCCATTATTTGGTCAAGTGGGACGATTAATAGTCCAGATA 180
 181 CGCCAGAGAGATCAAAATATTTGCTGTTATTTCAAGAGAGAGAAATTAATGCGGCG 240
 181 CGCCAGAGAGATCAAAATATTTGCTGTTATTTCAAGAGAGAGAAATTAATGCGGCG 240
 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGTATTTTGAAGCAACCGGCAATGCCAAC 300
 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGTATTTTGAAGCAACCGGCAATGCCAAC 300
 301 GATGCCAGATATTCGCAAGCGCTTACGCTAATAGTGAAGTATTAATCCAGAAAGTTCT 360
 301 GATGCCAGATATTCGCAAGCGCTTACGCTAATAGTGAAGTATTAATCCAGAAAGTTCT 360
 361 GGAATTAAGGCAATATTAACCAAGCTGACGAGAAACAGAGATTGATGACAGAA 420
 361 GGAATTAAGGCAATATTAACCAAGCTGACGAGAAACAGAGATTGATGACAGAA 420
 421 CAGTGCATATGCGTATTCGCTACCAAGCGCTAA 456
 421 CAGTGCATATGCGTATTCGCTACCAAGCGCTAA 456

RESULT 2
 AAC64620 standard; DNA; 456 BP.

AAC64620;
 26-FEB-2001 (first entry)
 Escherichia coli CsgB DNA sequence SEQ ID NO:4.
 Salmonella; agfa; Chromosomal gene replacement; fibrin; epitope;
 vaccine; immune response; immunogen; ds.
 Escherichia coli.
 MO200060102-A2.

PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 (UYVI-) UNIV VICTORIA.
 PA
 PI White AP, Doran JL, Collison SK, Kay MW;
 DR WPI: 2000-672631/65.
 XX P-ESDB; AAB36344.
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 134; 139p; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fibrillae (SEF17/TAPE) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fibrillae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;

Query Match 73.0%; Score 332.8; DB 3; Length 456;
 Best Local Similarity 83.1%; Pred. No. 1.5e-95; Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

1 ATGAAAAACAATTTGTTATTTATGATGTTGACAACTAGTGGTGGCTGGGATTGCAAC 60
 1 ATGAAAAACAATTTGTTATTTATGATGTTGACAACTAGTGGTGGCTGGGATTGCAAC 60
 61 GCGCAAAATTTGATCTGGCTGCTTGAATATTTTGGTAAATGAATTAAGCAAG 120
 61 GCGCAAAATTTGATCTGGCTGCTTGAATATTTTGGTAAATGAATTAAGCAAG 120
 121 TCTTCATTTATTCAGGGGGCCATTATTTGGTCAAGTGGGACGATTAATAGTCCAGATA 180
 121 TCTTCATTTATTCAGGGGGCCATTATTTGGTCAAGTGGGACGATTAATAGTCCAGATA 180
 181 CGCCAGAGAGATCAAAATATTTGCTGTTATTTCAAGAGAGAGAAATTAATGCGGCG 240
 181 CGCCAGAGAGATCAAAATATTTGCTGTTATTTCAAGAGAGAGAAATTAATGCGGCG 240
 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGTATTTTGAAGCAACCGGCAATGCCAAC 300
 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGTATTTTGAAGCAACCGGCAATGCCAAC 300
 301 GATGCCAGATATTCGCAAGCGCTTACGCTAATAGTGAAGTATTAATCCAGAAAGTTCT 360

```
Db 301 GATGCCATATTTGCGAAGTGTCTTATGTAATATCGATGATTTATCCAGAAAGTTCT 360
Oy 361 GGAATAATAGCCCAATATTATCCAGTACGCGAAGAAACAGCATGTTAGTGCAGAAA 420
Db 361 GGTATATAAAGCAATATTATACACAGTATGTACTCAAAAACGGCAATGTTAGTGCAGAGA 420
Oy 421 CAGTGCATATGCGTATGCGGTACCCCAAGCTTA 456
Db 421 CAGTGCATAATGCTATTGCGGTGACACAGCTTAA 456

RESULT 3
ACD68810
ID ACD68810 standard; DNA; 100 BP.
XX
AC ACD68810;
XX
ACD68810;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 80.
XX
KM Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EPI260592-AL.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI: 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX
PS Claim 3; Page 23; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population about gene expression. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SO Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;
Query Match 16.3%; Score 74.4; DB 7; Length 100;
Best Local Similarity 84.0%; Pred. No. 1.6e-13;
Matches 84; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
Oy 266 ACTTGCGTATATTGAGCAACGGGCAATGCCACGATCCAGTATATCCGAAGCGTT 325
Db 1 ACCTTGATATATTGATCAGCGGGCAGTGCACAGATGCAATTTCCGAAGGTGCTT 60
Oy 326 ACGGTAATAGTGCAGCTATTATTCAGAAAGTTCTGAGAA 365
Db 61 ATGGTAATACTGCGATGATTTATCCAGAAAGTTCTGGTAA 100

RESULT 4
ACD68812
ID ACD68812 standard; DNA; 100 BP.
XX
AC ACD68812;
XX
ACD68812;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 82.
XX
KM Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EPI260592-AL.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI: 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
PT patterns.
XX
PS Claim 3; Page 23; 2004pp; German.
XX
CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population about gene expression. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention
XX
SO Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;
Query Match 14.6%; Score 66.4; DB 7; Length 100;
Best Local Similarity 79.0%; Pred. No. 5.6e-11;
Matches 79; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Oy 123 TTCAATTAATCAGGCGGCATTTATTTGTCAAGTCCGACGAGTAAATAGTCCAGATGACG 182
```

The present invention describes a recombinant Agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA segment which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregated fibrillae (SEPI7/TAP) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fibrillae comprising recombinant Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) homologous recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier and introducing the useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. (1) is eliciting an immune response in an animal. In a fibrillar presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrillar protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrillar subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrillae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP: 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;


```

Query Match          8.2%; Score 37.6; DB 3; Length 456;
Best Local Similarity 51.2%; Pred. No. 0.15;
Matches      88; Conservative    0; Mismatches   84; Indels     0; Gaps       0

Oy      272 CGTATATTAGCAAAACGGGCAATGCACAACGATGTATTCGCCAAGAAGCGCTTAGCGTA 331
Db      203 CGACCATTATCCAGAGCGGTTATGTGAACGGGCGCATGTAGGCCAGSGTGCGGATAATA 262
Oy      332 ATAGTCAGCTATTATTCAGAAAAGGTTCTGAAATPAAGGCCAATATTACCCAGTACGTA 391
Db      263 GTACTATTGAACCTGCATCAGAAATGGTTTCAGAAAAATATATGCCACATCGACAGTGGAACG 322
Oy      392 CGCAAAAACAGCACTTGTAGTCAGAAAACACTCCGATATGGCTATTGCCGT 443
Db      323 CTAAAAACTCCGATATTACTGTCTGGCCCAATATGATCACGCTGGTACCCGTGT 374

RESULT 7
ADA71938/C
ID      ADA71938 standard; DNA; 2000 BP.
XX
XX      ADA71938;
XX
DT      20-NOV-2003 (first entry)
XX
DE      Rice gene, SEQ ID 5263.
XX
KW      Plant; bacterial infection; fungal infection; viral infection; rice;
XX      gene; ds.
XX
OS      Oryza sativa.
XX
MN      WO2003000898-A1.
XX
PN      03-JAN-2003.
XX
PF      22-JUN-2001; 2001WO-IB001105.
XX
PR      22-JUN-2001; 2001WO-IB001105.
XX
PP      22-JUN-2001; 2001WO-IB001105.
XX
PI      Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,
XX      Katagiri F, Quan S, Tao Y, Whitham S, Xie Z, Zhu T, Zhou G;
DR      MPI; 2003-175290/17.

PT      Identifying at least one gene involved in plant resistance or response to
PT      pathogenic infection for conferring resistance or tolerance to a plant to
PT      bacterial, fungal or viral infection by determining or detecting plant
PT      gene expression.
XX
PS      Claim 27; SEQ ID NO 5263; 89ppd; English.
XX
XX      The present invention relates to a method (M1) for identifying genes
XX      involved in plant resistance or response to pathogenic infection. M1
XX      comprises identifying a gene whose expression is significantly altered in
XX      the incompatible interaction of plant gene expression relative to
XX      expression of the gene in an uninfected plant, in a mutant plant that
XX      does not express a gene associated with response to pathogenic infection,
XX      or in a corresponding incompatible or compatible interaction. (M1) is
XX      useful for conferring resistance to resistance or tolerance to a plant to
XX      bacterial, fungal or viral infection. The present sequence was used to
XX      illustrate the invention.
SO      Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;

Query Match          8.2%; Score 37.6; DB 7; Length 2000;
Best Local Similarity 12.8%; Pred. No. 0.26;
Matches      57; Conservative 195; Mismatches 189; Indels    3; Gaps        2

Oy      14 TGTTATTATGATCTTGACATACCTACGCGGTCGCCCTGGGATTCGAACCGGACAAATTATNG 73

```

```

Db      1027 TGTAGTGCTCTKMKRRYRTYSMSFTYAMKKYTTTMTAYSSITWKMYAYVKRATYAMSSRKT 968
Qy      74  ATCTGGCTCGTTTCAGATAATAATTTTGCAGTAATGAATTAAGCAAGTCTTCATTATC 133
       ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      967  WWTGSGKEMATYCGTRKMAAAGRMRBMAMCWCMKKMKMTMSCKMKYRTWSCWYTMW 908
Qy      134  AGCGGGCCATTATTGTCAGATCGGCACGGATTAATGTGCCAGAGTAACCGAGAGAGAT 195
       ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      907  WGANRYAAYARRRRWTYKMSWRBYWTWTKAAWMTMTCMCAKMTMATGMAWMTMMWRYT 848
Qy      194  CAAACCTATTGTCGGTTATTTTCACAAAGAGGAAATAATCGGGGAAAGTGACACAG 253
       ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      847  MTTYTYANTCAKCKTKKMTATMTTWTACAMRITSRMRMAGMRKRYMKKRAYMRMRW 788
Qy      254  CAGGAAATTAATAACTTTGCGTATATTATTTAGCAAAACGGGCATGCCACGATGATAT 313
       ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      787  CMKGMAMMKMSRYMKMKKATATRYMKWMAATMMWSMRMKSRYRMWSGMRMRWSMRY 728
Qy      314  CGCAAAAGCGCTTACGTAATGACGACTATTATCAGAAAGTTTGGAAATAAGGCCA 373
       ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      727  CS--RMKCAKTKYASSARWTKRAKRSYRRRMYMRKMTYRRYWRSCMTFRAMSKR 670
Qy      374  ATATTACCGACGATCGGTACGCA--GAAACAGACAGTTGTATGTCAGAAACAGTCGATATG 432
       ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      669  RKMAGSMKSCMYWRRGRASMYSKYSGAKCCCKTRITWTSYSTGMTGMTSYKSMW 610
Qy      433  GCTATTTCGCGTACCCCAACGCTAA 456
       ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      609  TSKMSYMGKMTCTWYTSMKGSTRR 586
       ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 8
AAC64608
ID      AAC64608 standard; DNA; 39 BP.
XX      AAC64608;
AC      AAC64608;
XX      06-AUG-2003 (revised)
DT      26-FEB-2001 (first entry)
XX      XX
DE      Agfa (SEF17) recombinant agfa::P7ja generating PCR primer 17-A.
XX      XX
KW      Salmone11a; agfa; chromosomal gene replacement; fimbria; epitope;
KW      vaccine; immune response; immunogen; PCR primer; ss.
XX      XX
OS      Salmone11a enteritidis.
OS      Synthetic.
XX      XX
PN      WO200060102-A2.
XX      XX
PD      12-OCT-2000.
XX      XX
PE      05-APR-2000; 2000WO-CA000356.
XX      XX
PR      05-APR-1999; 99US-0127888P.
XX      XX
PA      (UYVI-) UNIV VICTORIA.
XX      XX
PI      White AP, Doran JL, Collison SK, Kay WW;
XX      WPI; 2000-672631/65.
XX      XX
PT      Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT      which encodes foreign epitope or antigen, expresses recombinant Agfa
PT      protein useful for eliciting immune response in animal.
XX      XX
PS      Example 2; Page 61; 13pp; English.
XX      XX
CC      The present invention describes a recombinant agfa gene (1) where a
CC      segment of the gene has been replaced by a segment of a foreign DNA
CC      sequence which encodes a foreign epitope or antigen. Also described are:
CC      (1) use of thin aggregative fimbriae (SEF17/TAR) nucleation depended
CC      assembly system of strains of Salmone11a, Escherichia coli and

```

CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
 CC correct OS field.)

SO Sequence 39 BP; 13 A; 7 C; 10 G; 9 T; 0 U; 0 Other;
 Query Match
 Best Local Similarity 7.9%; Score 35.8; DB 3; Length 39;
 Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 393 GCAGAAAAACAGCTGTAGTCAGAGAAACAGTGCCTAT 431
 Db 1 GCAGAAATTCAGCACTTGTAGTCAGAGAAACAGTGCCTAT 39

RESULT 9
 AAD32400
 ID AAD32400 standard; DNA; 738 BP.

AC AAD32400;
 DT 29-AUG-2003 (revised)
 DT 18-JUN-2002 (first entry)

DE Chlamydia pneumoniae nlpd gene.

KW Chlamydiaceae family; chronic infection; persistent infection; pyk; nlpd;
 KW Cpn0585; regulatory pathway; biosynthetic pathway; ompA; ompB; hsp60;
 KW lipopolysaccharide; cardiovascular system; respiratory tract; therapy;
 KW genital tract; reproductive system; atherosclerotic tissue; macrophage;
 KW multiple sclerosis; conjunctiva; prophylaxis; antibacterial; gene; ds.

OS Chlamydia pneumoniae.

FX Key Location/Qualifiers
 FT CDS 1..738
 FT /*tag= a
 FT /product= "Chlamydia pneumoniae nlpd protein"
 FT /note= "CDS does not include start codon"
 FT /partial

PN MO200214516-A1.

PD 21-FEB-2002.

PF 17-AUG-2001; 2001MO-AU001021.

PR 18-AUG-2000; 2000AU-00009540.

PA (YOU-) UNIV QUEENSLAND TECHNOLOGY.
 PA (MATH/) MATHEWS S A.

XX Timms P;
 XX
 XX

DR MPI; 2002-269197/31.
 DR P-PSDB; AAE20289.

PT Detecting Chlamydial organism in its persistent phase by detecting
 PT expression change of range of genes belonging to their respective
 PT biosynthetic pathways when expression is compared to that of organism in
 PT lytic phase.

XX Disclosure; Page 117-118; 196pp; English.

XX The invention relates to composition and methods for detecting organisms
 CC of the Chlamydiaceae family, including species of Chlamydia and
 CC Chlamydia, in the persistent phase of their developmental cycle and for
 CC the diagnosis of chronic or persistent infections caused by such
 CC organisms. The composition is useful for modulating the expression of
 CC genes such as pyk, nlpd, Cpn0585, a gene belonging to same regulatory/
 CC biosynthetic pathway and ompA, ompB, hsp60, a gene involved in
 CC lipopolysaccharide biosynthesis. It is also useful for modulating the
 CC level and/or functional activity of an expression product of these genes,
 CC where the gene is present in an epithelial cell (selected from
 CC cardiovascular system, respiratory tract, genital tract, reproductive
 CC system or conjunctiva), macrophage, or a cell associated with
 CC atherosclerotic tissue or associated with multiple sclerosis brain
 CC tissue. The composition is useful for treatment and/or prophylaxis of a
 CC chronic infection caused by an organism of the Chlamydiaceae family in a
 CC patient. Antigen associated with the persistent phase of the
 CC in the manufacture of a medicament, for treating and/or preventing
 CC Chlamydiaceae infection in a patient. The present sequence is Chlamydia
 CC pneumoniae nlpd gene. (Updated on 29-AUG-2003 to standardise OS field)

SO Sequence 738 BP; 231 A; 158 C; 166 G; 183 T; 0 U; 0 Other;

Query Match
 Best Local Similarity 7.7%; Score 35; DB 6; Length 738;
 Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 340 GCTATTATCCAGAAAGCTTCGAAATATAGCCATATTTCCAGTAGTCGACGAAA 399
 Db 1 GCTGTTGATCAGAAATATAGCAGGACGTAAATATGATGATGAGACATGTTAATA 60

OY 400 ACAGAGTGTAGTCGAGAAACAGTGCATATGCTATTCGCTCAC 446
 Db 61 ACAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 107

RESULT 10

ABK94407
 ID ABK94407 standard; DNA; 105325 BP.

AC ABK94407;

DT 27-AUG-2002 (first entry)

DE DNA encoding endothelin-3 (EDN-3).

KW Endothelin 3; EDN-3; endothelin converting enzyme; ECE;
 KW endothelin receptor; EDNR; signaling system; cardiovascular disease;
 KW coronary heart disease; hypertension; atherosclerosis; anglogenesis;
 KW fatty acid metabolism; diabetes; familial hypercholesterolemia;
 KW forensic marker; transgenic animal; solid support; SNP;
 KW cardiovascular regulator; gene; ds; single nucleotide polymorphism.

OS Homo sapiens.

FX Key Location/Qualifiers
 FT variation
 FT /*tag= a
 FT /replace(40215,A)
 FT /standard_name= "Single nucleotide polymorphism"
 FT /replace(59430,C)
 FT /*tag= b
 FT /standard_name= "Single nucleotide polymorphism"
 FT /replace(63843,T)

Bodian DL, Labow MA, Mickanin CS,

05-ÄPB-2000: 2000W0 0700000000

```
PR 05-APR-1999; 99US-0127888P.
XX
XX (UVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX
XX WPI: 2000-672631/65.
XX
XX P-PSDB; AAB36353.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;
XX
XX Query Match 7.5%; Score 34.2; DB 3; Length 456;
XX Best Local Similarity 56.8%; Pred. No. 1.8;
XX Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
XX
XX 275 ATATTGACCAAGCGGCAATGCCAGATATGCGAAGACGCTTAACGGTATA 334
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 206 AATGGCAGCATGCGATTATGTAACGGCGCGATGTAGGCCAGGTCCGATTAATGTA 265
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 335 GTGACGCTATTATCCAGAAAGTTCTGGAATTAAGCCCAATATTACCAGT 385
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 266 CTATTGAACCTACCTCAGAAATGTTTCAGAAATTAATGACCATGACCAAGT 316
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX RESULT 14
XX AAQ73066
XX ID AAQ73066 standard; DNA; 361 BP.
XX
XX AC AAQ73066;
XX
XX 27-AUG-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
XX
XX Salmonella, Agfa; vaccine, genetic immunization; ds.
XX
XX Salmonella enteritidis.
XX
```

```
FH Key Location/Qualifiers
FT CDS 1..359
FT /tag= a
FT /note= "Agfa"
FT misc_feature 37..60
FT /tag= d
FT /note= "TAF5 primer (pair with TAF6)"
FT 52..69
FT /tag= b
FT /note= "TAF3 primer (pair with TAF4)"
FT misc_feature complement(103..129)
FT /tag= e
FT /note= "TAF6 primer (pair with TAF5)"
FT complement(129..402)
FT misc_feature /tag= c
FT /note= "TAF4 primer (pair with TAF3)"
XX
XX W09425598-A2.
XX
XX 10-NOV-1994.
XX
XX 26-APR-1994; 94WO-IB000207.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX (KING/) KING J.
XX
XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI: 1994-358275/44.
XX P-PSDB; AAB62761.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
XX strains, vector constructs, or compens. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7A; 95pp; English.
XX
XX The DNA encodes the Salmonella enteritidis27655-3b TnpHoA mutant strain
XX agfa gene cloned into pUC19. The DNA and isolated proteins are used in
XX genetic immunization and vaccine compositions, respectively, to elicit an
XX immune response to Salmonella in animals (e.g. food producing animals)
XX and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
XX AUG-2003 to correct OS field.)
XX
XX Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
XX
XX Query Match 7.5%; Score 34; DB 2; Length 361;
XX Best Local Similarity 56.1%; Pred. No. 1.9;
XX Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
XX
XX 272 CGTATATTGACCAAGCGGCAATGCCAGATATGCGAAGACGCTTAACGGTA 331
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 140 CGACCATTAACCAAGCGGATGTAACGGCGCGATGTAGGCCAGGTCCGATTAATA 199
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Qy 332 ATAGTGACGCTATTATCCAGAAAGTTCTGGAATTAAGCCCAATATTACCAGT 385
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX Db 200 GTACTATTGAACCTACCTCAGAAATGTTTCAGAAATTAATGACCATGACCAAGT 253
XX | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX RESULT 15
XX AAT74141
XX ID AAT74141 standard; DNA; 361 BP.
XX
XX AC AAT74141;
XX
XX 25-MAR-2003 (revised)
XX DT 30-SEP-1997 (first entry)
XX
XX Salmonella enteritidis 27655-3b TnpHoA mutant agfa gene fragment.
XX
XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
```

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments) 5309.115 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaataacaatctgtatc.....ctcgccaccacacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgm2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgm2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgm2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgm2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgm2_6/ptodata/2/ina/PCUTS.COMB.seq: *
6: /cgm2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	41	9.0	7218	1	US-08-232-463-14
2	34	7.5	361	1	US-08-233-768A-56
3	34	7.5	456	1	US-08-233-768A-58
4	32.2	7.1	1443	4	US-09-328-352-2823
5	32.2	7.1	1830121	4	US-09-557-884-1
6	32.2	7.1	1830121	4	US-09-643-990A-1
7	32	7.0	651	4	US-09-134-000C-1323
8	31.8	7.0	1830121	4	US-09-557-884-1
9	31.8	7.0	1830121	4	US-09-643-990A-1
10	31.4	6.9	1230025	4	US-09-198-452A-1
11	31	6.8	1230025	4	US-09-198-452A-1
12	30.6	6.7	832	4	US-09-621-976-2813
13	30.6	6.7	1350	4	US-09-328-352-3386
14	30.2	6.6	1053	4	US-09-543-681A-408
15	30.2	6.6	1072	4	US-08-956-171E-435
16	30	6.6	1454	2	US-08-657-392-1
17	30	6.6	1454	2	PCT-US94-02539-1
18	30	6.6	1455	2	US-08-657-392-28
19	30	6.6	1455	2	PCT-US94-02539-26
20	30	6.6	9072	4	US-08-956-171E-45
21	30	6.6	92407	4	US-09-596-002-36
22	29.8	6.5	640681	4	US-09-790-988-1
23	29.6	6.5	474	4	US-09-621-976-18033
24	29.6	6.5	553	4	US-09-621-976-15491
25	29.6	6.5	1588	1	US-08-446-908-3
26	29.6	6.5	1588	1	US-08-231-205A-3
27	29.6	6.5	1588	2	US-08-871-161-3

ALIGNMENTS

28	29.6	6.5	1589	4	US-09-023-655-1132	Sequence 1132, Ap
C 29	29.4	6.4	141	4	US-08-956-171E-2326	Sequence 2326, Ap
C 30	29.4	6.4	501	4	US-08-956-171E-654	Sequence 654, App
31	29.4	6.4	1440	4	US-09-328-352-3130	Sequence 3130, Ap
32	29.4	6.4	1710	4	US-09-543-681A-934	Sequence 934, App
33	29.4	6.4	7253	4	US-09-268-347-35	Sequence 35, App1
34	29.4	6.4	9793	1	US-08-470-202-56	Sequence 56, App1
35	29.4	6.4	9793	1	US-08-471-770-56	Sequence 56, App1
36	29.4	6.4	9793	2	US-08-468-059-56	Sequence 56, App1
37	29.4	6.4	9793	3	US-09-109-916-56	Sequence 56, App1
38	29.4	6.4	9793	4	US-09-886-156-56	Sequence 56, App1
39	29.4	6.4	9793	4	US-09-886-149-56	Sequence 56, App1
40	29.4	6.4	9793	4	US-09-886-150-56	Sequence 56, App1
41	29.4	6.4	9793	4	US-09-886-159-56	Sequence 56, App1
42	29.2	6.4	1234	4	US-09-620-312D-953	Sequence 953, App
C 43	29.2	6.4	26284	4	US-09-526-193A-21	Sequence 21, App1
44	29	6.4	1287	4	US-09-461-325-18	Sequence 18, App1
45	29	6.4	1287	4	US-10-012-542-18	Sequence 18, App1

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHRIFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
CLASSIFICATION: 435
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO. 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpT-F18
US-08-232-463-14
Query Match 9.0%; Score 41; DB 1; Length 7218;

RESULT 6
 US-09-643-990A-1/c
 Sequence 1, Application US/09643990A
 Patent No. 6528289
 GENERAL INFORMATION:
 APPLICANT: Robert D. Fleischmann
 Mark D. Adams
 Owen White
 Hamilton O. Smith
 J. Craig Venter
 TITLE OF INVENTION: The Nucleotide sequence of
 the Haemophilus influenzae Rd Genome, Fragments
 Thereof, and Uses Thereof
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESS:
 ADDRESSER: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville,
 STATE: MD
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3 1/2 inch diskette
 COMPUTER: Dell Pentium
 OPERATING SYSTEM: MS DOS v6.22
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/643,990A
 FILING DATE: 23-AUG-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/487,429
 FILING DATE: 1995-06-07
 APPLICATION NUMBER: 08/426,787

Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 7.0% Score 31.8; DB 4; Length 1830121.
Best Local Similarity 52.7%; Pred. No. 34;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 7 AACAAATTGTTATGATGTTGACATACCTGGTGGCCCTGGGATTGCAACCGCGACA 66
Db 617140 AACACCTGTAAATTATGTTGTTCTCATATAGGTTGGATGCGACCGCCCGCGACA 617199

QY 67 AATTATGATCGGCTCGCTTCAGATATATATTTGGCGTAAATGAATTAGCAAGCTTCA 126
Db 617200 AATTACGATACCTGCTTTTCAAGCAACCATTTGACCGAGCAAAATTCAGTTT 617259

QY 127 TTTAATCGGC 137
Db 617260 TGTAAATGGGC 617270

RESULT 10
US-09-198-452A-1/c
Sequence 1, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849

SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (1) ..(15000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (15001) ..(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (30001) ..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (45001) ..(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (60001) ..(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (75001) ..(90000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (90001) ..(105000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (105001) ..(120000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (120001) ..(135000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (135001) ..(150000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (150001) ..(165000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (165001) ..(180000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (180001) ..(195000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (195001) ..(210000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (210001) ..(225000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (225001) ..(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (240001) ..(255000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (255001) ..(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (270001) ..(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (285001) ..(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (300001) ..(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (315001) ..(330000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (330001) ..(345000)

```

OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (345001)..(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (360001)..(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t

```

```

NAME/KEY: misc_feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature

```

Query Match
Best Local Similarity 56.2%; Score 31.4; DB 4; Length 1230025;
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

OY 68 ATTATGATCGCTCGCTGCAAGATTAATTTGGCGTAATGATTAAGCAAGTCTTCAT 127
DB 739726 ATCAAAATCTGATTTTTCAGGAATTAATGAGGCAAGTGGAAGTACCAACCTCTTGT 739667
OY 128 TTATAGGCGGCCATTATTTGTCAGAGTCGGCAGCATATATAGTG 172
DB 739666 TTGCAGATGAGGCGCATGCGCTTCAATTCATGATCAAGATTAATATG 739622

```

```

RESULT 11
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Glifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prever
; FILE REFERENCE: 9710-003-999
; CURRENT FILING DATE: US/09/198,452A
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA

```

1	OTHER INFORMATION: n=a or c or g or t
2	NAME/KEY: misc.feature
3	LOCATION: (360001)..(375000)
4	OTHER INFORMATION: n=a or c or g or t
5	NAME/KEY: misc.feature
6	LOCATION: (375001)..(390000)
7	OTHER INFORMATION: n=a or c or g or t
8	NAME/KEY: misc.feature
9	LOCATION: (390001)..(405000)
10	OTHER INFORMATION: n=a or c or g or t
11	NAME/KEY: misc.feature
12	LOCATION: (405001)..(420000)
13	OTHER INFORMATION: n=a or c or g or t
14	NAME/KEY: misc.feature
15	LOCATION: (420001)..(435000)
16	OTHER INFORMATION: n=a or c or g or t
17	NAME/KEY: misc.feature
18	LOCATION: (435001)..(450000)
19	OTHER INFORMATION: n=a or c or g or t
20	NAME/KEY: misc.feature
21	LOCATION: (450001)..(465000)
22	OTHER INFORMATION: n=a or c or g or t
23	NAME/KEY: misc.feature
24	LOCATION: (465001)..(480000)
25	OTHER INFORMATION: n=a or c or g or t
26	NAME/KEY: misc.feature
27	LOCATION: (480001)..(495000)
28	OTHER INFORMATION: n=a or c or g or t
29	NAME/KEY: misc.feature
30	LOCATION: (495001)..(510000)
31	OTHER INFORMATION: n=a or c or g or t
32	NAME/KEY: misc.feature
33	LOCATION: (510001)..(525000)
34	OTHER INFORMATION: n=a or c or g or t
35	NAME/KEY: misc.feature
36	LOCATION: (525001)..(540000)
37	OTHER INFORMATION: n=a or c or g or t
38	NAME/KEY: misc.feature
39	LOCATION: (540001)..(555000)
40	OTHER INFORMATION: n=a or c or g or t
41	NAME/KEY: misc.feature
42	LOCATION: (555001)..(570000)
43	OTHER INFORMATION: n=a or c or g or t
44	NAME/KEY: misc.feature
45	LOCATION: (570001)..(585000)
46	OTHER INFORMATION: n=a or c or g or t
47	NAME/KEY: misc.feature
48	LOCATION: (585001)..(600000)
49	OTHER INFORMATION: n=a or c or g or t
50	NAME/KEY: misc.feature
51	LOCATION: (600001)..(615000)
52	OTHER INFORMATION: n=a or c or g or t
53	NAME/KEY: misc.feature
54	LOCATION: (615001)..(630000)
55	OTHER INFORMATION: n=a or c or g or t
56	NAME/KEY: misc.feature
57	LOCATION: (630001)..(645000)
58	OTHER INFORMATION: n=a or c or g or t
59	NAME/KEY: misc.feature
60	LOCATION: (645001)..(660000)
61	OTHER INFORMATION: n=a or c or g or t
62	NAME/KEY: misc.feature
63	LOCATION: (660001)..(675000)
64	OTHER INFORMATION: n=a or c or g or t
65	NAME/KEY: misc.feature
66	LOCATION: (675001)..(690000)
67	OTHER INFORMATION: n=a or c or g or t
68	NAME/KEY: misc.feature
69	LOCATION: (690001)..(705000)
70	OTHER INFORMATION: n=a or c or g or t
71	NAME/KEY: misc.feature
72	LOCATION: (705001)..(720000)
73	OTHER INFORMATION: n=a or c or g or t

```
NAME/KEY: misc feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
```

```
Query Match
Best Local Similarity 6.8%; Score 31; DB 4; Length 1230025;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
Qy 344 TTATCCGAAGAGGTTCTGGAATAAGGCCAATATTATCCGATGCGTACGACGAGAAACAG 403
Db 1044632 TGATCAGAAAGATGCGAGCGTGAATATGATCGTAGAGCATGTGTAATACAG 1044691
Qy 404 CAGTTGATGTCAGAAAAGTCGATATGCTATTCGCGTAC 446
Db 1044692 CTGTCGATGTAATGCTATGCTTTCGTCAC 1044734
```

```
RESULT 12
US-09-621-976-2813
Sequence 2813, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 2813
LENGTH: 832
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
```

```
NAME/KEY: CDS
LOCATION: 235..399
US-09-621-976-2813
```

```
Query Match
Best Local Similarity 6.7%; Score 30.6; DB 4; Length 832;
Matches 21; Conservative 92; Mismatches 76; Indels 0; Gaps 0;
Qy 10 AAATGTTATTTATGATGTTGACATCTGGCGCTGGATGCAACCGACCAAT 69
Db 40 WRRKKKAAWKKWTWYRMAWMTYKKAACRTKTKKKKWWMMWYWGWRSSYM 99
Qy 70 TATGATCGGCTGTCAGAAATATTTTGGCGTAATGAATTAAGCAAGTCTCATTT 129
Db 100 AAWTRTWTGYAYRSMWYRRCWKAKYRKTCTSSKWTWKKKAAWTTWMMKT 159
Qy 130 AATCAGCGCGCATTTATGTCAGAGTCGACGACGATATAGTGCAGAGTACCGAGNA 189
Db 160 YYMAATRYMMWMTWTRWASWYCMWGGKAKWSTWRSRSYASARSACRCSCSWGA 219
Qy 190 GGATCAAAA 198
Db 220 MSWKYMRM 228
```

```
RESULT 13
US-09-328-352-3386/c
Sequence 3386, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 3386
LENGTH: 1350
ORGANISM: Acinetobacter baumannii
US-09-328-352-3386
```

```
Query Match
Best Local Similarity 6.7%; Score 30.6; DB 4; Length 1350;
Matches 45; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
Qy 309 TATATCGCAAGCGGTTACGTAATAGTACAGCTATATCCAGAAAGTTCTGAAATTA 368
Db 586 TATATGAAAAGCTGTTTGAGCTAATGCCAGTATTTCCATTAAAGCTTGGAATAA 527
Qy 369 GGCCCAAT 377
Db 526 TGCCACATT 518
```

```
RESULT 14
US-09-543-681A-408/c
Sequence 408, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 408
LENGTH: 1053
TYPE: DNA
```

ORGANISM: Proteus mirabilis
US-09-543-681A-408

Query Match 6.6%; Score 30.2; DB 4; Length 1053;
Best Local Similarity 62.7%; Pred. No. 3;
Matches 47; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

OY 322 GCTTACGGTATAGCAGCATTTATCCAGAAAGTCTGGAATAAGCCCATTTTACC 381
DB 366 GCCTTCATTACATTACAGATTATCAAGAAAGATCAGAAATAGGCTTTCACACC 307
OY 382 CAGTACGGTACGACG 396
DB 306 AATAACGGTACGCG 292

RESULT 15
US-08-956-171E-435

Sequence 435, Application US/08956171E
Patent No. 6593114

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

Gil H. Choi

Patrick S. Dillon

Craig A. Rosen

Steven C. Barash

Michael R. Farnon

TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5256

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/956,171E

FILING DATE: 20-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/009,861

FILING DATE: January 5, 1996

APPLICATION NUMBER: 08/781,986

FILING DATE: January 3, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman

REGISTRATION NUMBER: 46,789

REFERENCE/DOCKET NUMBER: PB248P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (240) 314-1224

TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 435:

SEQUENCE CHARACTERISTICS:

LENGTH: 1072 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 435:

US-08-956-171E-435

Query Match 6.6%; Score 30.2; DB 4; Length 1072;
Best Local Similarity 49.3%; Pred. No. 3;
Matches 68; Conservative 3; Mismatches 67; Indels 0; Gaps 0;

OY 98 TTGGGTAAATGAATTAAGCAAGTCTTCATTTAATCAGGCGGCATTATTGTCAGTGC 157
DB 6 TTTTGTAARGAGAAATCAATATGAACAAGTATAAGCGTATTTAATCATTTAGATG 65

OY 158 GCAAGATAATATAGTCCAGAGTACGCCAGAAAGATCAAACTATTGTCCGTTATTTCAC 217
DB 66 GCACAAATGTATATATGGAAACAGATGAGATTGTGAGCAANCAATTCATTCATTATTTA 125
OY 218 AAGAAGGAGAAATATATC 235
DB 126 ATGTAAAGGCAATTCCTC 143

Search completed: March 16, 2004, 04:36:29
Job time : 62.6647 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million Cell updates/sec

Title: US-09-543-407-2

Perfect score: 456

Sequence: 1 atgaaacaaatcgtatc.....ctcgccaccccaacgctaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubna/US09C_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubna/US09C_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	7.7	738	US-10-369-435-3	Sequence 3, Appl1
2	34.4	7.5	18540	US-10-231-416-3	Sequence 3, Appl1
3	34.4	7.5	18540	US-10-231-426-3	Sequence 3, Appl1
4	34	7.5	529	US-10-027-632-241642	Sequence 241642,
5	33.8	7.4	203654	US-09-820-905-3	Sequence 3, Appl1
6	33.4	7.3	598	US-10-027-632-206661	Sequence 206661,
7	33.4	7.3	598	US-10-027-632-206662	Sequence 206662,
8	33.4	7.3	598	US-10-027-632-206663	Sequence 206663,
9	33.4	7.3	660	US-10-369-493-29323	Sequence 29323, A
10	33.2	7.3	535	US-10-449-857A-18	Sequence 18, Appl1
11	33.2	7.3	1444	US-10-425-114-1093	Sequence 1093, Ap
12	32.8	7.2	2649	US-10-094-749-1500	Sequence 1500, Ap
13	32.4	7.1	1914	US-09-815-242-4167	Sequence 4167, Ap
14	32.4	7.1	1929	US-09-815-242-8045	Sequence 8045, Ap
15	32.2	7.1	1163020	US-10-398-221-1	Sequence 10, Appl1

C 16	32.2	7.1	1830121	14	US-10-329-960-1	Sequence 1, Appl1
C 17	32.2	7.1	1830121	15	US-10-329-670-1	Sequence 1, Appl1
C 18	32.2	7.1	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
C 19	32	7.0	849	15	US-10-027-632-130879	Sequence 130879,
C 20	32	7.0	849	15	US-10-027-632-130880	Sequence 130880,
C 21	32	7.0	849	15	US-10-027-632-130881	Sequence 130881,
C 22	32	7.0	1353	12	US-10-282-122A-21063	Sequence 21063, A
C 23	32	7.0	1356	9	US-09-815-242-6450	Sequence 6450, Ap
C 24	32	7.0	1539	12	US-10-282-122A-25267	Sequence 25267, A
C 25	32	7.0	7640	12	US-10-424-599-51912	Sequence 51912, A
C 26	32	7.0	7640	9	US-09-070-927A-375	Sequence 375, App
C 27	31.8	7.0	583	15	US-10-027-632-14367	Sequence 14367, A
C 28	31.8	7.0	933	9	US-09-815-242-6967	Sequence 6967, Ap
C 29	31.8	7.0	933	12	US-10-282-122A-22047	Sequence 22047, A
C 30	31.8	7.0	1365	9	US-09-815-242-4243	Sequence 4243, Ap
C 31	31.8	7.0	1365	9	US-09-815-242-4196	Sequence 4196, Ap
C 32	31.8	7.0	2209	9	US-09-925-301-494	Sequence 494, App
C 33	31.8	7.0	5599	12	US-10-363-616-89	Sequence 89, Appl1
C 34	31.8	7.0	1830121	14	US-10-329-960-1	Sequence 1, Appl1
C 35	31.8	7.0	1830121	15	US-10-329-670-1	Sequence 1, Appl1
C 36	31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, Ap
C 37	31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, Ap
C 38	31.6	6.9	1251	15	US-10-369-493-44921	Sequence 44921, A
C 39	31.6	6.9	465237	9	US-09-933-267A-1	Sequence 1, Appl1
C 40	31.4	6.9	472	9	US-09-864-761-2780	Sequence 2780, Ap
C 41	31.4	6.9	664	15	US-10-027-632-135259	Sequence 135259,
C 42	31.4	6.9	664	15	US-10-027-632-135260	Sequence 135260,
C 43	31.4	6.9	664	15	US-10-027-632-152059	Sequence 152059,
C 44	31.4	6.9	808	15	US-10-027-632-160710	Sequence 160710,
C 45	31.4	6.9	808	15	US-10-027-632-160711	Sequence 160711,

ALIGNMENTS

RESULT 1

US-10-369-435-3

Sequence 3, Application US/10369435

Publication No. US20040002440A1

GENERAL INFORMATION:

APPLICANT: Mathews, Sarah

FILE REFERENCE: 10338-15US (2615070/VPA)

CURRENT FILING DATE: 2003-02-19

PRIOR APPLICATION NUMBER: US/10/369,435

PRIOR FILING DATE: 2000-08-18

PRIOR APPLICATION NUMBER: PCT/AU01/01021

NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn version 3.2

SEQ ID NO 3

LENGTH: 738

TYPE: DNA

ORGANISM: Chlamydia pneumoniae

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(738)

US-10-369-435-3

Query Match

Best Local Similarity 57.9%; Pred. No. 1.9;

Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

DB 340 GCTATTATCCAGAAAGCTTCGAAATAGGCCATATTACCCAGTACGATCCAGAA 399

DB 1 GCTTTATATCGAAGAAATGACGACGAGTGAATATGATCTAGAGATGTATATA 60

DB 400 ACACAGTTGTATGTCGAGAAACAGTCCGATATGCTATTCGCGTAC 446

DB 61 ACACCTGTCGATGAAATGATATATGCTTGTGCTCTTTTCGTAC 107

CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 203654
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(203654)
OTHER INFORMATION: n = A,T,C or G
US-09-820-905-3

Query Match 7.4% Score 33.8; DB 9; Length 203654;
Best Local Similarity 50.3%; Pred. No. 86;
Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

Qy 5 AAAACAAATGTTATTTATGATGTTGACAAATACCTGGCGCTGGAGTTGCAACCGCGA 64
Db 71389 AGAATATTTTCTTTTCTTTTGTAGATGTTTACTGAAAACCTAGCGAACAACATGCT 71330
Qy 65 CAAATATGATCTGCTGCTGCTGCAATATTAATTTGCGTAATGATTAAGCAAGCTTT 124
Db 71329 CAACGAACTAATGCTTCTCTGTAATATGTTTGAAGAAAGTATTAACCCCAAT 71270
Qy 125 CATTAATCAGCGCGCATTAATGCTCAAGTCGCGCAGATTAAT 169
Db 71269 ACTAAGAACAGGAACCAAGATGAGATGTGGAACATGAAAT 71225

RESULT 6
US-10-027-632-206661/c
Sequence 206661, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 206661
LENGTH: 598
TYPE: DNA
ORGANISM: Human
US-10-027-632-206661

Query Match 7.3% Score 33.4; DB 15; Length 598;
Best Local Similarity 51.7%; Pred. No. 5.5;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 200 TATGTCCGTTATTTCAACAAGAGAAATATCGGGCGAAGTGCACGAGCAGGGA 259
Db 563 TATATTTCTTAATTCATAGAGATCTTTAGAAATTAACCTTAACAATTAAGATACATT 504
Qy 260 ATTAACTTTCGTATATTGAGCAACGGGCAATGCGCAAGATGCCAGTATATCGCAA 319

Db 503 TTTAAATGTGCATTTATAGGAGCTGATAAATGCTAATTAATATAGCATTTAGTAAT 444
Qy 320 GCGCTTACGTTAATATAGTCACCTATTA 346
Db 443 GCTTAATTAATAGCATTGAGAGTGTGA 417

RESULT 7
US-10-027-632-206662/c
Sequence 206662, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 206662
LENGTH: 598
TYPE: DNA
ORGANISM: Human
US-10-027-632-206662

Query Match 7.3% Score 33.4; DB 15; Length 598;
Best Local Similarity 51.7%; Pred. No. 5.5;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
Qy 200 TATGTCCGTTATTTCAACAAGAGAAATATCGGGCGAAGTGCACGAGCAGGGA 259
Db 563 TATATTTCTTAATTCATAGAGATCTTTAGAAATTAACCTTAACAATTAAGATACATT 504
Qy 260 ATTAACTTTCGTATATTGAGCAACGGGCAATGCGCAAGATGCCAGTATATCGCAA 319
Db 503 TTTAAATGTGCATTTATAGGAGCTGATAAATGCTAATTAATATAGCATTTAGTAAT 444
Qy 320 GCGCTTACGTTAATATAGTCACCTATTA 346
Db 443 GCTTAATTAATAGCATTGAGAGTGTGA 417

RESULT 8
US-10-027-632-206663/c
Sequence 206663, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483

QY	200	TATGTCCGTTATTTTCCAAAGAGGAAATATTCGGCGGAAGTCACACGCGAGGA	259
Db	563	TATATCTTTTATTCATAGAGATCTTGTAAAGATTAGCCTAAACAATTAGATACAT	504
QY	260	ATTATACTTTCGGATATTAGCAAAACGGGCATGCAACGATGCCATATATGCGAA	319
Db	503	TTTTTAATGTGCATTATTAGCGAGTGTATAATGTATATTATACATTTAGTAT	444
QY	320	GGCCTTACGTAATAGTCAGACTTTA	346
Db	443	GCTTTAATATATAGCATTTGAGGTCTGA	417

	Query March Best Local Matches	Similarity 55.7%; 64;	Score 33.4; Conservative	DB 15;	Length 660;	Pred. No. 5,8;	Mismatches 51;	Indels 0;	Gaps 0
QY	250	CAGCGCAGGAATTATTA	CTTGCCTATTTTGA	CAACGCGCAATGCGCAATGCGCAAT	309				
Db	339	CAATCGGCAATTTGCCG	ATTGGCAGATTGGCC	AGACCAATTCGCGCCATTCCTGA	280				
QY	310	ATATGCGCAAGGCGTT	AGCGTAATATGTCG	AGCTAATTAACGAAAGGTCTGGAA	364				
Db	279	ATTATTAGTGCATCTAC	AGTAATATCTTTAT	AGTATCATCATCAGGCTCTGGAA	225				

```

1: Sequence ID, Application US/10449857A
2: Publication No. US20040043931A1
3: GENERAL INFORMATION:
4: APPLICANT: Hersberg, Robert M.
5: APPLICANT: Hosken, Nancy A.
6: APPLICANT: Lodes, Michael J.
7: APPLICANT: Mohamath, Rashed
8: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
9: TITLE OF INVENTION: AND DIAGNOSIS OF INFLAMMATORY BOWEL DISEASE
10: FILE REFERENCE: 584C1
11: CURRENT APPLICATION NUMBER: US/10/449, 857A
12: CURRENT FILING DATE: 2003-05-30
13: NUMBER OF SEQ ID NOS: 89
14: SOFTWARE: Corixa Invention Disclosure Database
15: SEQ ID NO 18
16: LENGTH: 535
17: TYPE: DNA
18: ORGANISM: Unknown
19: FEATURE:
20: OTHER INFORMATION: Unknown Bacterium
US-10-449-857A-18

```

Query Match	7.3%;	Score 33.2;	DB 12;	Length 535;
Best Local Similarity	51.3%;	Pred. No. 6;		
Matches	77;	Conservative	0;	Mismatches 73.

[illegible]

```

RESULT 11
US-10-425-114-3093
; Sequence 3093, Application US/10425114
; Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jirong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 3093
LENGTH: 1444
TYPE: DNA
FEATURE:
ORGANISM: Zea mays
OTHER INFORMATION: Clone ID: 700240986_FLI
US-10-425-114-3093

```

Query Match	7.3%	Score 33.2	DB 12	Length 1444
Best Local Similarity	55.1%	Pred. No. 10		
Matches	65	Conservative 0	Mismatches 53	Indels 0
				Gaps 0
QY	249	CCAGCAGAGAAATATACTTGGTATATTGCGAAACGGCGCATGCCACGATGCCAG	308	
Db	207	CCATGGAATTATATACACCCACTTACGCAAGAAACAGTGGGAGAAAATCTTCCTG	266	
QY	309	TATATCGAAGCGCTTACGTAATATGTCAGACTTATCCAGAAAGTTCTCGAAAT	366	

Db 267 CTCAGCGTACAGTGTCTCCCTGTTCTAGTGAATATTTCCAGGAAGCTTGCAATT 324

RESULT 12
US-10-094-749-1500
; Sequence 1500, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: MAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT FILING DATE: 2002-03-12
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1500
; LENGTH: 2649
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1500

Query Match 7.2%; Score 32.8; DB 15; Length 2649;
Best Local Similarity 48.4%; Pred. No. 19;
Matches 91; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

Qy 208 GTTATTTCAAGAAAGAGAAATATTCGGCGAAAGTCAGCAGGAGGATTTATAC 267
Db 807 GTTCTTATCTAGCAGAGAGATACAAAAGCTCATAGTCAGTCCAAAGCTTCACTAC 866
Qy 268 TTTCGGTATATTGAGCAACGGGCAATGCCAAGTCCAGTATATCGCAAGGCGTTAC 327
Db 867 CTGGTCTGCTGAGTGAAGTCTGTCATATCCCTGGAAAGCATATGCAAGTATCAGTAAA 926
Qy 328 GGTATATGTGAGCTATATTCAGAAAGCTTCTGAAATAGGCCAATATTTACCAAGTAC 387
Db 927 AATATCTGCGCTGGGAAAAATCCAGAAATATCCAGAGAAATGATACCTAAACCAAGAC 986
Qy 388 GGTACGCA 395
Db 987 TTAACGCA 994

RESULT 13
US-09-815-242-4167
; Sequence 4167, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4167
; LENGTH: 1914
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4167

Query Match 7.1%; Score 32.4; DB 9; Length 1914;
Best Local Similarity 51.4%; Pred. No. 21;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 263 ATAAGTTGGTATATGAGAAACGGCAATGCCAAGTCCAGTATATTCGAAGCG 322
Db 1001 AAAATTGAAATTTGTTATCAAACTGCAATTTACCAACTATATAGATAGGTCTCTA 1060
Qy 323 CTACGGTAAATAGTACAGCTATATTCAGAAAGTTCTGAAATATAGGCCAATTTATACC 382
Db 1061 AAGCGATCATATATGCAATCATTTGGCCAAATGGTATTTGAAATATCACCCTTAATTTAAA 1120
Qy 383 AATACGATACCAAGAAACGAGTT 408
Db 1121 CTAATGCTAATCAACAAAAGCACTT 1146

RESULT 14
US-09-815-242-8045
; Sequence 8045, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 8045
;; LENGTH: 1929
;; TYPE: DNA
;; ORGANISM: Staphylococcus aureus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1929)
US-09-815-242-8045

Query Match
Best Local Similarity 7.1%; Score 32.4; DB 9; Length 1929;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 263 ATAACTTGCCTATTATGAGCAAGGCGCATGCGCATATCGCAAGCG 322
DB 1001 AAAATTGGAATTCGTTATAACTGCAATTACCAACCTATCAGTATAGAGTCTCTA 1060
QY 323 CTACGGTAATAGTGCAGCTATTATCCGAAAGGTTCTGGAATAGGCCATATTACC 382
DB 1061 AAGCGATCATATAGCATATTCGGCCAAATGTTATGAAATGACCTTAATTAAAA 1120
QY 383 AGTACGGTACGCAAGAAACGCGAGTT 408
DB 1121 CTATTGCTAATCAACAAAGCACTT 1146

RESULT 15

US-10-398-221-10
;; Sequence 10, Application US//10398221
;; Publication No. US20040018514A1
;; GENERAL INFORMATION:
;; APPLICANT: KUNST, Frederick
;; APPLICANT: GLASER, Philippe
;; TITLE OF INVENTION: *Listeria innocua*, genome and applications
;; FILE REFERENCE: 344 702 - US
;; CURRENT APPLICATION NUMBER: US/10/398,221
;; PRIOR FILING DATE: 2003-03-27
;; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;; PRIOR FILING DATE: 2001-10-04
;; PRIOR APPLICATION NUMBER: FR 00/12 697
;; NUMBER OF SEQ ID NOS: 4025
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 10
;; LENGTH: 1163020
;; TYPE: DNA
;; ORGANISM: *Listeria innocua*
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(end)
;; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

Query Match
Best Local Similarity 7.1%; Score 32.2; DB 15; Length 1163020;
Matches 108; Conservative 0; Mismatches 108; Indels 1; Gaps 1;

QY 215 CACAAGAGAGAAATATATCGGCGCAAGTCGACGAGGAGGAATTAACCTTGGCT 274
DB 1136649 CACAAGACATAGAGAAATTCGTAATTCGTTAGATTAATAAATCTATTAACTG 1136708
QY 275 ATATTGACAAATCGGCGCATGCCAAGATCCAGATATATCGAAAGCGCTTACGTAATA 334
DB 1136709 ATATTGAAGTACGGGCGTATTGTAAATGAAGATAGATGTCAGAGAAATTAATGAGATA 1136768

QY 335 GTGACGCTATTATCCAGAAAGCTTCTGAAATTAAGGCCAATATTACCAGTACGTAACG 394
DB 1136769 GT-TAGCGAGATATATAGTATGCTTAAAGAAAGATATGAAATTTAAATATAGCAATGAA 1136827
QY 395 AGAAACAGCAGGTTGTAGTGCAGAAACAGTCGCATAT 431
DB 1136828 CGCTACAAACCAATTTACAGTATGAAACTCAGGAATTT 1136864

Search completed: March 17, 2004, 08:15:08
Job time : 410.736 secs

Email: est@atcon.wustl.edu
Library constructed by Bruce Blumberg
DNA Sequencing by: Washington University Genome Sequencing Center
Source lab clone id - 3749963
Trace considered overall poor quality This clone is available
(image@image.jnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1..220

FEATURES
source
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3749963"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 P"
/clone_lib="Xenopus laevis gastrula non normalized"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2 ug of poly A+ RNA (equal parts from stage 10.5 and stage 11.5 gastrulae). EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. The library was mass excised and used to infect Top10F. Clones were picked into freezing medium (per liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4 7 H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 24 hours. Original library construction by Bruce Blumberg (Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Query Match 29.0%; Score 132.2; DB 12; Length 220;
Best Local Similarity 75.6%; Pred. No. 5.2e-30;
Matches 164; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY CTGCTTCAAGATTAATTTTGGCTAAATGATTAAGCAAGCTTCAATTAATCAGGCG 139
DB 3 CTAAATTCATATTAATTAATCTGGCGGCAATGATGAGTCTTCAATTAATCAGGCG 62
QY 140 CCATTATTTGTCAGTGGCAGGATTAATAGTCAGAGTACGCGCAGGAAGATCAAAAC 139
DB 63 CCAATATTTGTCAGTGGCAGGCTAATTAATAGTCTCAAGAGGAGGCTCAAAAC 122
QY 200 TATTCCTGCTTATTCACAAGAGAGAAATATTCGCGCAAGTGCAGCAGGAGGGA 259
DB 123 TTTTGGCGTCTCGCAAGAGGTATGACCAACCGGCGAAGATTGACACAGCAGAG 182
QY 260 ATTATTAATCTTGGTATATTTAGCAAAAGGCGCAATGC 296
DB 183 ATTATTAATCTTGCATATATGATCAGCGGCGGCTCC 219

RESULT 2
A1368352 207 bp mRNA linear EST 11-JAN-1999
LOCUS A1368352/c
DEFINITION SMOVMFCAR04G09SK Onchocerca volvulus microfilaria cDNA
(SAM98MLW-OWMF) Onchocerca volvulus cDNA clone SMOVMFCAR04G09 5',
mRNA sequence.
ACCESSION A1368352
VERSION A1368352.1 GI:4147105
KEYWORDS EST.
SOURCE Onchocerca volvulus
ORGANISM Onchocerca volvulus
Bukariyota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Onchocerca.
REFERENCE 1 (bases 1 to 207)
AUTHORS Williams, S. A.
JOURNAL Genes expressed in microfilaria of Onchocerca volvulus
COMMENT Unpublished (1999)
Contact: Steven A. Williams
Molecular Parasitology

Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 413/5853826
Fax: 413/5853786
Email: genome@smith.edu
Seq primer: pBluescript SK.
Location/Qualifiers
1..207

FEATURES
source
/organism="Onchocerca volvulus"
/mol_type="mRNA"
/db_xref="taxon:6282"
/clone="SMOVMFCAR04G09"
/dev_stage="microfilaria"
/lab_host="XLI-Blue MRF"
/clone_lib="Onchocerca volvulus microfilaria cDNA
(SAM98MLW-OWMF)"
/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10⁵ independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr. S.A. Williams, email: genome@smith.edu."

ORIGIN

Query Match 11.8%; Score 53.6; DB 9; Length 207;
Best Local Similarity 91.8%; Pred. No. 1.4e-05;
Matches 56; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATGAAAACAATGCTATTTATGATTTGACAAATCTGGTGGCTGGATTGCAACC 60
DB 61 ATGAAACCCNATTTGATTTATGATGATTAACAATCTGGTGGCTGGATTGCAACC 2
QY 61 G 61
DB 1 G 1

RESULT 3
CNS00E93 1101 bp DNA linear GSS 04-JUN-1999
LOCUS CNS00E93
DEFINITION Drosophila melanogaster genome survey sequence TERT end of BAC #
BACR28G06 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL068795
VERSION AL068795.1 GI:4949039
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Bukariyota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS
JOURNAL
COMMENT Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamonos in Pictet de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain y2; cn bw sp, the same strain used for the BACPac p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPac Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

Location/Qualifiers
1. .1101

```

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR28G06"
/clone_lib="RPCI-98"
/note="end : TET3"

```

ORIGIN

	Query Match	8.6%	Score 39.4	DB 29	Length 1101
	Best Local Similarity 33.2%		Pred. No. 0.61		
	Matches 92	Conservative 52	Mismatches 133	Indels 0	Gaps 0
Qy	2	TGAAAAA	CAAAATTGTTATTATGATGTTGACAAATCTGGTGCCTGGAGATTGCAACCG	61	
Db	825	KGAAAAAA	AAACCTTGATTAATGATGAAATGMAAAASKGTITCTCTCTTTTMDTTGGGS	884	
Qy	62	CGACAAAT	TATGATCTGCTCGTTGAGATATTAATTTGGCGTAATGATTAATGACAAAGT	121	
Db	885	TGCMTTAA	TATATYKKGCGGAAGRGKTAARPDTSGDAKTRITATTSATSMETMAATTARB	944	
Qy	122	CTTCATT	TAAATCAGCGCGCAATTATGTCGAATCGGACGGAATATAGTCCACAGTAC	181	
Db	945	KSTAAAT	TKSGSCCAAMCCCTTTTTTBTTTTTTTTTKGVGGAAAMFTTKTKTKTGT	1004	
Qy	182	GCCAGAAG	AGATCAAAACTATGTCCGTTATTTTCACAGAAGAGGAATATATCGGCGCA	241	
Db	1005	TKKARGA	GCGGCGGAAAAATTTGGAGTCMAATDABARARACGRARRARARARGRGA	1064	
Qy	242	AAATCGAC	CGGCGGGAATTAATACCTTGGCTATAT	278	
Db	1065	AAAMRGAA	AAAGATDAAMTKGTMTTTTGGRTGAAWAT	1101	

FEATURES

```
1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
```

```
/clone="XL186b22"  
/tissue_type="whole embryo"  
/dev_stage="stage 10.5"  
/clone_lib="N1BB Mochizuki normalized Xenopus early gastrula  
library"
```

ORIGIN

Query Match 8.6%; Score 39; DB 12; Length 660;
 Best Local Similarity 89.4%; Pred. No. 0.65;
 Matches 42; Conservative 0; Mismatches 5; Indels 0;
 Gaps 0;
 QY 410 TAGTGCAGAAACAGTGCATATGCTATTGCGCTACCAACGCTAA 456
 680 TAGTGCAGAGACAGTGCCTAAATGCTATTGCGCGTACCAACGCTTAA 634

FEATURES

```

1..775
/organism="Trypanosoma cruzi"
/mol_type="mRNA"
/strain="CI - Brenner"
/db_xref="taxon:5693"
/clone="n1059.r"
/cell_type="epimastigote"
/clone_id="T. CruzI epimastigote normalized cdna library"
/note="Site 1: SCor1; Site 2: Ncrl; cDNA library
constructed with oligo dt primed epimastigote mRNA
and modified polyinifer"
cloned in pUC138d phagemid with modified polyinifer"

```

ORIGIN

Query Match	8.5%;	Score 38.8;	DB 9;	Length 775;
Best Local Similarity	57.4%;	Pred. No. 0.84;		
Matches 70;	Conservative 0;	Mismatches 52;	Indels 0;	Gaps 0;

AZ578608/c 288 bp DNA linear GSS 08-DEC-2000
 LOCUS 25d10 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
 DEFINITION
 ACCESION SP. NGR234 genomic clone 25d10, genomic survey sequence.
 VERSION
 AZ578608.1 GI:11606060
 KEYWORDS
 SOURCE
 ORGANISM
 Rhizobium sp. NGR234
 Rhizobium sp. NGR234
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 1 (bases 1 to 288)
 Virey, V., Rosenthal, A., Broughton, W. J. and Perret, X.
 Genetic snapshots of the Rhizobium species NGR234 genome
 Genome Biol. 1 (6), RESEARCH0014 (2000)
 2114532
 11178268
 11178268
 CONTACT: Virginie Virey
 Laboratoire de Biologie Moléculaire des Plantes Supérieures
 University of Geneva
 1 Chemin de l'Impératrice, Chambesey/Geneva 1292, Switzerland
 Tel: +44(0)1603450000
 Fax: +44(0)1603450045
 Email: virginie.virey@bsrc.ac.uk
 Class: shotgun.
 FEATURES
 source
 1..288
 /organism="Rhizobium sp. NGR234"
 /mol_type="genomic DNA"
 /strain="ANU265"
 /db_xref="taxon:394"
 /clone="25d10"
 /note="Shot-gun genomic library of Rhizobium strain
 ANU265"
 /note="Vector: M13; derivative strain of NGR234 cured of
 pNGR234a"

Query Match 8.3%; Score 38; DB 28; Length 288;
 Best Local Similarity 50.0%; Pred. No. 1.1;
 Matches 95; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

ORIGIN
 204 GTTCGTTATTTCACAGAGAGAAATATATGCGGCGAAATGACACGAGGAGAAATTA 283
 238 GTCCGTCGTCTTTCAGAGAGAGAAATATATCTCCGAGCGGCGATTCGGAAGA 179
 264 TAACCTTGCGTATTTAGCAAGCGGCAATGCGCAAGATGATATCGCAAAAGCG 323
 178 CCAAGAGCGATCACCTCGCAAAAGGCGAATGACAGCTTCGGGCACTTCAGTTCCG 119
 324 TTACGCTAATAGTGCAGCTATTATTCAGAAAGTTTGAATAAGGCCAATTTACCA 383
 118 TTACGCGCAACAATCGTGAACCCACAGAGAGAAATATCTCTCCATCAGAGTGA 59
 384 GTACGCTAGC 393
 58 GGACGATCG 49

RESULT 7
 BP007561 675 bp mRNA linear EST 15-MAR-2002
 LOCUS BP007561 Nori Satoh unpublished cDNA library, young adult Ciona
 DEFINITION
 ACCESION intestinalis cDNA clone ciad45a18 5', mRNA sequence.
 VERSION
 BP007561.1 GI:19499038
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Ciona intestinalis
 Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Ciona.
 1 (bases 1 to 675)
 Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.

TITLE
 JOURNAL
 COMMENT
 Expressed genes in Ciona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 FEATURES
 source
 1..675
 Location/Qualifiers
 1..715
 /organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="ciad45a18"
 /tissue_type="whole animal"
 /dev_stage="young adult"
 /clone_lib="Nori Satoh unpublished cDNA library, young
 adult"

Query Match 8.3%; Score 37.8; DB 12; Length 675;
 Best Local Similarity 50.3%; Pred. No. 1.6;
 Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

ORIGIN
 229 AATATCGGCGCAAGTGCACGAGGAGATTAATTAATTTGGCTATTTAGCAACG 288
 257 AATCTTCAATGAAAGAAATGTCGAAACTGTAAGTAAAGGAAAGAA 316
 289 GGCATGCGCAAGATGCGCATATATCCCAAGCGCTTACGTAATAGCTATTAATC 348
 317 TCCCTCACCTCGCAGGCAATGATATGATCAAAAGAAACAAATTAAGCTTGAACCT 376
 349 CAGAAAGTTCTGGAATTAAGCCCAATTAATCCAGTACGTAAGCAAGAAACAGCAGTT 408
 377 CAATTAATTTCTGCAATTAATGACAAATTCACAGATGATTAACACACAGAT 436
 409 GTAGT 413
 437 GAACCT 441

RESULT 8
 AV896333/c 715 bp mRNA linear EST 09-NOV-2001
 LOCUS AV896333 Nori Satoh unpublished cDNA library, young adult Ciona
 DEFINITION
 ACCESION intestinalis cDNA clone ciad45a18 3', mRNA sequence.
 VERSION
 AV896333.1 GI:16885431
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Ciona intestinalis
 Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Ciona.
 1 (bases 1 to 715)
 Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.
 Expressed genes in Ciona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 FEATURES
 source
 1..715
 Location/Qualifiers
 1..715
 /organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="ciad45a18"
 /tissue_type="whole animal"
 /dev_stage="young adult"

QY 131 ATCAGCGCCATATATGCTCAATCGGACGCGATTAATGTCAGAGTACGCCAGAG 190
 DB 967 ATCCAGGGAACTTTATATGATGAGAGAACCATCCGTTAATTCCTGCTCTTGA 908
 QY 191 GATCAAACTATATGCTCGTATTTTCAAGAGAGGAAATTAATGGGCGGAAAGTCC 250
 DB 907 TAGCAAAATGGATACCTTAAGAGAACAAAGAACTGGCCATTGCGGTCTCCAA 848
 QY 251 AGGCAAGGATTAATACCTTGGCTATATTGAGCAAAAGCGGCAATGCCAGATGCCAGT 309
 DB 847 TACCGAGAAACAATATTTATCTATCTGCTCAACAGCTTCAAGAGATTAATCTAT 789

RESULT 11
 BU387712 708 bp mRNA linear EST 28-NOV-2002
 LOCUS 603857374F1 CSEQCHN75 Gallus gallus cDNA clone CHEST863c20 5', mRNA
 DEFINITION
 ACCESSION BU387712
 VERSION BU387712.1 GI:25895713
 KEYWORDS EST,
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus.
 1 (bases 1 to 708)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 22335534
 1245392

COMMENT
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk.

FEATURES
 source
 1..708
 Location/Qualifiers
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST863c20"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_lib="CSEQCHN75"
 /note="Organ: trunks; Vector: pBluescript II KS(+); Site_1: EcoRI; Site_2: NotI; This normalized library was constructed from 1 million independent clones. cDNA synthesis was initiated using an oligo(dT) primer, using methylated C in the first strand synthesis reaction. Following this first strand reaction, double-stranded cDNA was blunt-ended, ligated to NotI adapters, digested with EcoRI, size-selected, and cloned into the NotI and EcoRI compatible sites of a custom modified MCS of the pBluescript (KS+) vector. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6 (1996): 791, except that a significantly longer reannealing hybridization was used."

ORIGIN
 Query Match 8.1%; Score 36.8; DB 13; Length 708;
 Best Local Similarity 50.6%; Pred. No. 3.4;
 Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 89 AATATATTTTGGCGTAATGAATTAAGCAAGTCTCATTTAATCAAGGCGGCATTATTG 148

DB 17 AAGAGATCTACCCGTAATATGACAGACACAGCTGAGACTGACTGAGGCTCTATCA 76
 QY 149 GTCAGTCGACGAGATTAATAGTCCAGAGTACGACGAGAGGATCAAAATATGTCG 208
 DB 77 GTCCCAAGCCAAAGAAAGGCTTGGCCACTGATTTCAGAAAGAGTGAAGAGGTGCCA 136
 QY 209 TTAATTCACAAAGAGAGAAATATATCGGCGAAGTGCACGACGAGGAAATTA 264
 DB 137 GTCTCTTCTGATGACGTAATATTAAGTGAAGGCTATAGAACGGCAATTTT 192

RESULT 12
 A2529672 924 bp DNA linear GSS 03-NOV-2000
 LOCUS ENTBM10TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION
 ACCESSION A2529672
 VERSION A2529672.1 GI:11082756
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 924)
 Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000)
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@igr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seg primer: M3-Reverse
 Class: Shotgun
 High quality sequence start: 33
 High quality sequence stop: 745.
 Location/Qualifiers
 1..924
 Location/Qualifiers
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

ORIGIN
 Query Match 8.1%; Score 36.8; DB 28; Length 924;
 Best Local Similarity 50.6%; Pred. No. 3.7;
 Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 73 GATCGGCTCGTCAATATATTTGGCGTAATGAATTAACAGTCTTCAATTAAT 132
 DB 740 GATGTACACTTAAGAGTTTATTAAGGGGTAAACATTTAGGGGTATTACATTGAA 799
 QY 133 CAGCGGCAATTAATGCTCAAGTGGACGAGTATATAGTGCACAGTACGACAGAGA 192

Db 800 AAGACATATTGTCATCTGTCATGGTACGGGAGGAGATACACCATATGA 859

Qy 193 TCAAACTATGTCGTTATTTCAAGAGAGAAATATCGGCGCAAGTCA 248

Db 860 TGTAAAGAAATGTCAGTATGATTTGAGAGAGGAGTTAAACGCGACAGACA 915

RESULT 13

CC211905/1550 bp DNA linear GSS 12-MAY-2003

LOCUS CH261-18606.sp6.1 CH261 Gallus gallus genomic clone CH261-18606,

DEFINITION genomic survey sequence.

ACCESSION CC211905

VERSION CC211905.1 GI:30530573

KEYWORDS GSS

SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus

REFERENCE 1 (bases 1 to 1550)

AUTHORS Kremetzki, C., Higgindocham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.

TITLE Gallus gallus BAC End Reads

JOURNAL Unpublished (2003)

COMMENT Contact: Richard K. Wilson

Genome Sequencing Center

Washington University School of Medicine

Email: submissions@wustl.edu

Insert Length: 18200 Std Error: 0.00

Seq primer: Sp6 ATTTAGGTGACACTATAG

Class: BAC ends

High quality sequence start: 339

High quality sequence stop: 1106.

Location/Qualifiers

1..1550

/organism="Gallus gallus"

/mol_type="genomic DNA"

/strain="Red Jungle Fowl"

/db_xref="taxon:9031"

/clone="CH261-18606"

/sex="Female"

/cell_line="UCD001, inbred 256"

/clone_lib="CH261"

/note="Vector: pTRABAC2.1; Site 1: EcoRI, Site 2: EcoRI; CH261 Female Chicken library - For library and clone ordering information: <http://www.choxi.org/bacpac>"

ORIGIN

Query Match 8.0%; Score 36.4; DB 28; Length 1550;

Best Local Similarity 59.8%; Pred. No. 5.9;

Matches 61; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

Qy 97 TTTCGGTAATGATTAAGACGCTTCATTATATCGGCGCATTTGTCATGTC 156

Db 1185 TTTCGGTGTGATGATTAAGCTGCTGTTCTTGTATATGGAAGACAATTTCTCTC 1126

Qy 157 GCGACGATTAATGATGCGAGATGACGCGAGAGAGATCAAA 198

Db 1125 AGCTCAGGAATATCTCTGTAGTTCAGGGGAGGAGAAAAA 1084

RESULT 14

BH528896/499 bp DNA linear GSS 14-DEC-2001

LOCUS BOGPO61TR BOCP Brassica oleracea genomic clone BOGPO61, genomic

DEFINITION survey sequence.

ACCESSION BH528896

VERSION BH528896.1 GI:17744694

KEYWORDS GSS

SOURCE Brassica oleracea

ORGANISM Brassica oleracea

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.

1 (bases 1 to 499)

AUTHORS Town, C.D., Van Aken, S., Uterback, T., Koo, H. and Frazer, C.M.

TITLE Whole genome shotgun sequencing of Brassica oleracea

JOURNAL Unpublished (2001)

COMMENT Other GSSs: BOGPO61TF

Contact: Chris Town

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523

Fax: 301-838-0208

Email: cdtown@tigr.org

DNA is from a doubled haploid provided by Tom Osborn.

Seq primer: TR

Class: sheared ends.

Location/Qualifiers

1..499

/organism="Brassica oleracea"

/mol_type="genomic DNA"

/strain="T01000DH3"

/db_xref="taxon:3712"

/clone="BOGPO61"

/clone_lib="BOGP"

/note="Vector: pHOS1, Site 1: BstXI, 2-3 kb sheared genomic DNA inserted into pHOS1 using BstXI linkers"

ORIGIN

Query Match 7.9%; Score 36; DB 28; Length 499;

Best Local Similarity 46.4%; Pred. No. 5.4;

Matches 117; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 183 CCAGAGAGATCAAAACTATGTCGTTATTTCAAGAGAGAGAAATATCGGCGAA 242

Db 470 CCGGAAACGAGCAAAATCATTCACGATGATGAAGATGACAGGGCAATGCTATCA 411

Qy 243 AGTCGACGAGCGGAGGAAATATTAATCTTGCTATATGAGAAACGGGCAATGCCACGA 302

Db 410 AGAAGATCAATGAGGAGAGATGATGATTTTCATGAGCTATTCAGACTATTTGA 351

Qy 303 TGCCAGTATTCGCAAGCCCTTACGCTTAATAGTACGATTTATTCAGAAAGGTTCTG 362

Db 350 GGGGACAGAGAGAAACCGTAATGCGTCTCTATCAACAATGTGAGCGGAGAAAGAA 291

Qy 363 AATAAGCCCAATTTAACCAAGTACGCTGACGAGAAACGCACTTGTATGCGAAACA 422

Db 290 CGAGAACTATTAACGCTCCAAAGCTGACAGATCAAGAGGCGCTATTTCTATTTCA 231

Qy 423 GTCCATATGGC 434

Db 230 TGCCGACAAAGC 219

RESULT 15

BG226083/579 bp mRNA linear EST 09-MAY-2001

LOCUS KQ08h12.y1 TB95TM-SSR Strongyloides stercoralis cDNA 5' similar to

DEFINITION WP:2K686.3 CE00457, mRNA sequence.

ACCESSION BG226083

VERSION BG226083.1 GI:12713638

KEYWORDS EST

SOURCE Strongyloides stercoralis

ORGANISM Strongyloides stercoralis

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Panagrolaimidae; Strongyloidae; Strongyloides.

1 (bases 1 to 579)

REFERENCE McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritzer, E., Bennett, J., Franklin, C., Tsagaris, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter, JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH (nutman@nih.gov). DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
High quality sequence stop: 417.

FEATURES

1..579
Location/Qualifiers
/organism="Strongyloides stercoralis"
/mol_type="mRNA"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/lab_host="XL-1 Blue MRP" (Stratagene)"
/clone_lib="TBN95FM-SSR"
/note="Vector: Lambda Uni-ZAP XR (Stratagene), Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
rhabditiform larvae which had been isolated from gerbils
experimentally infected with larvae originally isolated
from experimentally infected dogs. cDNA was constructed
and, using adaptors, was cloned unidirectionally into the
vector from the EcoRI site to the XhoI site. The library
has an unamplified titer of 1 x 10E5 pfu/ml and an
amplified, undiluted titer of 9 x 10E11 pfu/ml. The
average insert size of the unamplified library is 675 bp
(range, 100-1700)."

ORIGIN

Query Match 7.9%; Score 36; DB 12; Length 579;
Best Local Similarity 49.0%; Pred. No. 5.6;
Matches 96; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

QY 7 AACAAATTGTATTATGATGTCACAATGCTGGGCTGGGATTGCAACCGCACA 66
   |||||
Db 271 AAAAAACTTATTGTTATGTTGATTATGACAGATGCTCCAGATTTTGCTTCATA 330
   |||||
QY 67 AATTATGATCTGGCTGCTTCAGAAATATATTTTGGGTAATGATTAAGCAAGCTTCA 126
   |||||
Db 331 AATCTTAATACAGCTCCCGCATTTATACATTTTCCAGCTAAGGATCTAGGAATCTGAT 390
   |||||
QY 127 TTTAATCAGGGCGCATTTATGTCAGTCGCGCAGATTAATGATGCGCAGTACGCCAG 186
   |||||
Db 391 GATTCAAATGATGTTCAAAATTAATGATTTGAAGCTGATGTTTGGCAAGATTGTTAAA 450
   |||||
QY 187 GAAGATCAAAACTAT 202
   |||||
Db 451 GAAGAAGCTGAAATAT 466
   |||||

```

Search completed: March 16, 2004, 04:28:38
Job time : 2238.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 seconds
(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaacttttaaaagtaga.....cgaccgtctacagctactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_ets: *
12: gb_gy: *
13: gb_un: *
14: gb_vt: *
15: gb_ba: *
16: gb_fun: *
17: gb_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_scs: *
28: em_un: *
29: em_vt: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rtd: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_gy: *
39: em_hcgo_hum: *
40: em_hcgo_mus: *
41: em_hcgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	456	100.0	456	6	AX814811	AX814811 Sequence
2	456	100.0	648	1	ECOCGAA	L04979 Escherichia
3	454.4	99.6	4680	1	ECSCGABDG	X90754 E.coli csrg
4	454.4	99.6	10346	1	AE000205	AE000205 Escherich
5	454.4	99.6	15047	1	D90741	D90741 Escherichia
6	395	86.6	1711	1	AE275733	AE275733 Escherich
7	395	86.6	10190	1	AE005315	AE005315 Escherich
8	395	86.6	327773	1	AP002554	AP002554 Escherich
9	382.2	83.8	306358	1	AE016759	AE016759 Escherich
10	340.8	74.7	10370	1	AE015131	AE015131 Shigella
11	340.8	74.7	292504	1	AE016981	AE016981 Shigella
12	311.4	68.3	437	1	AF237726	AF237726 Shigella
13	259.2	56.8	5103	1	STAJ2301	AU002301 Salmonell
14	259.2	56.8	22411	1	AE008749	AE008749 Salmonell
15	257.6	56.5	254050	1	AL627269	AL627269 Salmonell
16	257.6	56.5	301983	1	AE016840	AE016840 Salmonell
17	254.4	55.8	456	6	I44909	I44909 Sequence 58
18	254.4	55.8	2067	1	SEU43280	U43280 Salmonella
19	253.6	55.6	1048	1	STAGFBA	AU000514 Salmonell
20	252.8	55.4	2883	1	ESAS15702	AU515702 Enterobac
21	246.4	54.0	2889	1	CSP515700	AU515700 Citrobact
22	230.8	50.6	2920	1	CPR515701	AU515701 Citrobact
23	186	40.8	19201	1	D90742	D90742 Escherichia
24	173.4	38.0	361	6	I44908	I44908 Sequence 56
25	121.8	26.7	230	1	SEU53207	U53207 Salmonella
26	83.8	18.4	1212	1	EC0131756	AU131756 Escherich
27	78	17.1	78	6	AX814809	AX814809 Sequence
28	72	15.8	72	6	AX814798	AX814798 Sequence
29	66	14.5	66	6	AX814808	AX814808 Sequence
30	43.4	9.5	301214	1	AE016786	AE016786 Pseudomon
31	43.2	9.3	85600	9	AC138647	AC138647 Homo sapi
32	42.6	9.3	135259	9	HS127B20	Z83838 Human DNA B
33	42.6	9.3	165074	9	AC093519	AC093519 Homo sapi
34	41.6	9.1	181047	2	AC016765	AC016765 Homo sapi
35	41.6	9.1	198599	9	AC108448	AC108448 Homo sapi
36	41.2	9.0	159070	2	AC128114	AC128114 Rattus no
37	41.2	9.0	186562	2	AL451079	AL451079 Homo sapi
38	41.2	9.0	200368	9	AC096541	AC096541 Homo sapi
39	41.2	9.0	241394	2	AC132789	AC132789 Rattus no
40	40.8	8.9	165074	9	AC093519	AC093519 Homo sapi
41	40.6	8.9	236533	2	AC103433	AC103433 Rattus no
42	40.2	8.8	183458	2	AC129554	AC129554 Mus muscu
43	40	8.8	180995	9	AC117500	AC117500 Homo sapi
44	40	8.8	187615	2	AC020724	AC020724 Homo sapi
45	39.8	8.7	999	1	AY077581	AY077581 Bradyrhiz

ALIGNMENTS

RESULT 1
AX814811
LOCUS AX814811 456 bp DNA
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Bacteri: Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Bioerck,L., Olsen,A., Wikstroem,M. and Herwald,H.
Peptides
Patent: WO 03064446-A 15 07-AUG-2003;
JOURNAL

FEATURES

Location/Qualifiers (SE)

1..456
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..456
/note="unnamed protein product"

CDS

/codon_start=1
/transl_table=11
/protein_id="CAE83541.1"
/db_xref="GI:39104002"
/translation="MKLKYKALIVFSSGALAVPQYGGGNGHGGGNSPSE
LDQNGKNSBMTVKQFGGNGAAYDQTASNSVAVTVQVFGNNATAHQY"

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 456; DB 6; Length 456;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTAGAGCAATTCGACCAATCGATTCTCCGGTAGCGCTTGGA 60
1 ATGAACTTTTAAAGTAGAGCAATTCGACCAATCGATTCTCCGGTAGCGCTTGGA 60
61 GGTGTTCTCTCACTACGCGCGCGCGGTAAACACCGGTGATGAGCGGTAATAGCGG 120
61 GGTGTTCTCTCACTACGCGCGCGCGGTAAACACCGGTGATGAGCGGTAATAGCGG 120
121 CCAATTCTGAGCTGAACATTACCAATGAGCGGTGATGAGCGGTAATAGCGGTAAT 180
121 CCAATTCTGAGCTGAACATTACCAATGAGCGGTGATGAGCGGTAATAGCGGTAAT 180
181 ACTGATGCGCGTAATCTGACTGATTAATCCAGCATGCGCGGTGATGAGCGGTAAT 240
181 ACTGATGCGCGTAATCTGACTGATTAATCCAGCATGCGCGGTGATGAGCGGTAAT 240
241 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATGAGCGGTAATAGCGGTAATAG 240
241 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATGAGCGGTAATAGCGGTAATAG 240
241 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATGAGCGGTAATAGCGGTAATAG 300
301 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 360
301 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 360
361 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 360
361 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 360
361 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 420
421 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 420
421 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 456
421 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 456

RESULT 2

LOCUS

ECOCSSGA

DEFINITION

Escherichia coli curlin subunit (csgA) gene, complete cds.

VERSION

L04979.1 GI:290424

KEYWORDS

csgA gene; curli organelle; fibronectin-binding protein.

SOURCE

Escherichia coli

ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Escherichia.

1 (bases 1 to 648)

The Rpos sigma factor relieves H-NS-mediated transcriptional

in Escherichia coli

JOURNAL

MOL. Microbiol. 7 (4), 523-536 (1993)

MEDLINE

93211294

PUBMED

8459772

COMMENT

On Jun 11, 1993 this sequence version replaced gi:145630.
Original source text: Escherichia coli (sub_strain W3110, strain
K-12) (library: Kohara) DNA.
Location/Qualifiers

FEATURES

source

1..648
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/issue="11b"
/gene="csgA"
83..538
/gene="csgA"
83..538
/note="precursor protein"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="curlin subunit"
/protein_id="AAA23616.1"
/db_xref="GI:290425"
/translation="MKLKYKALIVFSSGALAVPQYGGGNGHGGGNSPSE
LDQNGKNSBMTVKQFGGNGAAYDQTASNSVAVTVQVFGNNATAHQY"

CDS

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 456; DB 1; Length 648;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

sig_peptide
mat_peptide

/evidence=experimental
/gene="csgA"
143..535
/product="curlin subunit"
/evidence=experimental

1 ATGAACTTTTAAAGTAGAGCAATTCGACCAATCGATTCTCCGGTAGCGCTTGGA 60
83 ATGAACTTTTAAAGTAGAGCAATTCGACCAATCGATTCTCCGGTAGCGCTTGGA 142
61 GGTGTTCTCTCACTACGCGCGCGGTAAACACCGGTGATGAGCGGTAATAGCGG 120
143 GGTGTTCTCTCACTACGCGCGCGGTAAACACCGGTGATGAGCGGTAATAGCGG 202
121 CCAATTCTGAGCTGAACATTACCAATGAGCGGTGATGAGCGGTAATAGCGGTAAT 180
203 CCAATTCTGAGCTGAACATTACCAATGAGCGGTGATGAGCGGTAATAGCGGTAAT 262
181 ACTGATGCGCGTAATCTGACTGATTAATCCAGCATGCGCGGTGATGAGCGGTAAT 240
263 ACTGATGCGCGTAATCTGACTGATTAATCCAGCATGCGCGGTGATGAGCGGTAAT 322
241 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATGAGCGGTAATAGCGGTAATAG 300
323 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATGAGCGGTAATAGCGGTAATAG 382
301 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 360
383 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 442
361 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 420
443 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 502
421 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 456
503 GGTACTCTGATCAGTGAAGCAAGCAAAATTTCTGAATGAGCGGTAATAGCGGTAAT 538

RESULT 3

ECCSGABDG

gene
CDS
sig_peptide
gene
CDS
misc_feature
misc_feature
misc_feature
CDS
gene
CDS
gene
CDS
sig_peptide
gene
CDS

AQAQNSVYDPEBYNDPFGIETSPSALDNFQAIQSGIILGGLSNINTGKPRGMWTDNYIV
DIANDDQGLQNVNVDKRTQTQVSGLQNNSTDP"
/complement(1434. .1823)
/gene="cs98"
/complement(1434. .1823)
/gene="cs98"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curl1
production"
/protein_id="CA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translation="MKRYLRMIVAAEFLFAAGNLHAVEVPGLLTDHTVSSIGHDFEY
RAPSFKMSDYGNLTITNERSAPKGSWITTVQDVYFQFFLPPLKRDPEKTVFAL
IOTELALRRIOINALLSTGDLAHDEF"
/complement(1828. .2478)
/gene="cs9D"
/complement(1828. .2478)
/gene="cs9D"
/note="necessary for transcription of the cs9A operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
/protein_id="CA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translation="MFNEVHSIRHGTLLLTTKSSQLATALLQHLKQSLATGKLNHQ
KSLDDISSGSIILDMEMADKGLIHVQDILSRKNNIKITLLTPEDYPRDIENP
HINGFPYSMEQERVNGLOGVLRGECYFTQKLASYLIHSGNTRYNSTSALITHEE
KEILNKRIKRGASNEIEARSLISENTVTLVNLFKKIAYVNRQAVSMANDNLR"
/complement(2626)
/note="transcriptional start site for cs9DRC operon"
3140
/note="transcriptional start site for cs9A operon"
3233. .3688
/gene="cs9B"
3233. .3688
/gene="cs9B"
/codon_start=1
/transl_table=1
/product="nucleation component of curlin monomers"
/protein_id="CA62281.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translation="MKKNLLFPMULTLIGAPGIAAAGYDIANSEYNFAVELSKSFNN
QAALIQAGTNNSSKROGSGKLVAVAOESSNRAKITDGTGVNLAIVDOAGSANDA
SISQAVYNTMTIIOKSGKNKANTYGTGKTALIVQROSMAIRVQR"
3729. .4184
/gene="cs9A"
3729. .4184
/gene="cs9A"
/note="major subunit"
/codon_start=1
/transl_table=1
/product="curlin/Cs9A protein"
/protein_id="CA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="MKLLKVAIAIAIVFSSAGSLAGVVPYGGGGNGGGGNSGNSPE
LNTIYGGGNSALALQTDPARSDLTTHRGGNCADVDGQSDDSIDLTHRGGNSAT
LDQNNGNKSEITVAFQFGGNGAAYDTPANSSVNVTVGFGNNATAHQY"
3729. .3788
/gene="cs9A"
4243. .4575
/gene="orfC"
4243. .4575
/gene="orfC"
/codon_start=1
/transl_table=1

/product="putative curli production protein"
 /protein_id="CA62283.1"
 /db_xref="GI:1147565"
 /translation="MTLLLAASSLTPTTQGGVYTIPEVTLTOSCLCRVQL
 SLREGSSQSOTKQEKRTSLPANQPIALTKLSLNPDRKIVIVTSDQSILHSQ
 WPSSEKS"

ORIGIN

Query Match 99.6%; Score 454.4; DB 1; Length 4680;
 Best Local Similarity 99.8%; Pred. No. 3,5e-121;
 Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGTAGAACATTCGACCAATTCGATTCCTCGGTACCGCTTGGCA 60
 DB 3729 ATGAACTTTAAAGTAGAACATTCGACCAATTCGATTCCTCGGTACCGCTTGGCA 3788
 QY 61 GGTGTTTCTCTCACTACGCGCGCGCGGTAAACACAGGTGTGGCGGTATATATACCGGC 120
 DB 3789 GGTGTTTCTCTCACTACGCGCGCGCGGTAAACACAGGTGTGGCGGTATATATACCGGC 3848
 QY 121 CCMAATTCGAGCTGAACATTTACCACTAGCGGTGCGGTACTCTGCACCTTGTGCA 180
 DB 3849 CCMAATTCGAGCTGAACATTTACCACTAGCGGTGCGGTACTCTGCACCTTGTGCA 3908
 QY 181 ACTGATGCCGTAACCTCTGACTTGAATTAACCAAGCGCGGTATATGTCACAT 240
 DB 3909 ACTGATGCCGTAACCTCTGACTTGAATTAACCAAGCGCGGTATATGTCACAT 3968
 QY 241 GTTGTCAGGCGCTAGATGACCTCAATGATCAATCCCAACGTCGCTCGGTAAACAGC 300
 DB 3969 GTTGTCAGGCGCTAGATGACCTCAATGATCAATCCCAACGTCGCTCGGTAAACAGC 4028
 QY 301 GCTACTCTTATATAGTGAACCGCAAAATTCGAAATAGACGCTTAAACAGTTCGCGT 360
 DB 4029 GCTACTCTTATATAGTGAACCGCAAAATTCGAAATAGACGCTTAAACAGTTCGCGT 4088
 QY 361 GCGAACGCGTCTCAGTTCACCAAGACTGATCTTAATCTCTCGTCAACGTACTCAGTT 420
 DB 4089 GCGAACGCGTCTCAGTTCACCAAGACTGATCTTAATCTCTCGTCAACGTACTCAGTT 4148
 QY 421 GCGTTTGTGAACACGCGACCGCTCATGACTACTAA 456
 DB 4149 GCGTTTGTGAACACGCGACCGCTCATGACTACTAA 4184

RESULT 4
 AEO00205 10346 bp DNA linear BCT 01-DEC-2000
 LOCUS Escherichia coli K12 MG1655 section 95 of 400 of the complete
 DEFINITION genome.
 ACCESSION AEO00205 U00096
 VERSION AEO00205.1 GI:1787265
 KEYWORDS
 SOURCE
 ORGANISM Escherichia coli K12
 Escherichia coli K12
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 10346)
 Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mu, B. and Shao, Y.
 The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)
 JOURNAL MEDLINE 9742617
 PUBMED 9278503
 2 (bases 1 to 10346)
 Blattner, F.R.
 Direct Submission
 TITLE Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 E-mail: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:

REFERENCE 608-263-7459
 3 (bases 1 to 10346)
 Authors
 Title
 Journal
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 E-mail: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:

REFERENCE 4 (bases 1 to 10346)
 Authors
 Title
 Journal
 Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source
 Location/Qualifiers
 1..10346
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 complement(1..48)
 /note="IS3"

repeat_region
 promoter
 promoter
 promoter
 5..32
 /note="Factor Sigma70; predicted +1 start at 1094689"

gene
 69..392
 /gene="b1028"
 69..392
 /gene="b1028"
 /function="orf, Unknown"

note="0107; This 107 aa ORF is 27 pct identical (0 gaps)
 to 58 residues of an approx. 1888 aa protein PAS2_CANAL
 SW: P43098"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC74113.1"
 /db_xref="GI:1787266"
 /translation="MEDGVQAMRDYLAGLDIASPEHOVLNVTAKSEVAPSLIKENTLS
 LHLTHYKWTESLDTPLMPTPVAFLFISNPKYLGMLNDFAGVDDQRVWRCKKAFSD
 AKVFK"
 275..299
 /gene="b1028"
 /note="Factor Sigma32; predicted +1 start at 1094983"

promoter
 gene

/gene="ycdu"
/note="synonym: b1029"
389. .1375
/gene="ycdu"
/function="orf; Unknown"
/note="0328; This 328 aa ORF is 19 pct identical (4 gaps)
to 191 residues of an approx. 288 aa protein Y320_MYCGE
SW: P47562"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"
/translation="MIDPVLTFIRFODKRNLIYIYALIGLIGPYMKAGFPFSEDI
GVSGIATLVYKNIYFDLKAAYAKCYATKRNIDPSFKKKQKHKEFLTPVAGFSL
LIMASWGLIQLPSLYALFLISLGLPVIPLFRNIRTSYKQVIAVAKVKYK
SLTRYVLSVCSITVNLITISPLNDSFVTEGWLTFKSIIMALLIGCVLAINLF
FLRFSKRYAFLGRFLQIDILFSESNALSTFPAKPLMLRLFILIVEMNITVSVL
ATLVEMRIWPEAYFLCYVPCLIYFYFCFLMNDPMACDMYERMGHPNK"
1253. .1280
/gene="ycdu"
/note="central position to predicted promoter: -104"
/bound_molecly="DeOR predicted site"
1253. .1280
/gene="ycdu"
/note="No predicted promoter"
/bound_molecly="DeOR predicted site"
1337. .1365
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1096049"
1494. .1745
/gene="b1030"
1494. .1745
/gene="b1030"
/function="orf; Unknown"
/note="083; This 83 aa ORF is 25 pct identical (7 gaps)
to 81 residues of an approx. 616 aa protein ALBU_RABIT SW:
P49065"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGSTRVOYTKKKPVLSYELFKYGGEGGIDLSRSCGPANHC
VHGSSNLSPEVDRSGILLIPVCAIHEKKARTFRALL"
1513. .1926
/gene="ycdu"
/note="synonym: b1031"
1513. .1926
/gene="ycdu"
/function="putative structure; Not classified"
/note="0137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein RS37_YEAST SW:
P05759"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKKSYPFRTSSLIAMAVRGGLTFAPAPAGSPLTAFVCP
GROLSTPGRFSFSPRCVQYTKKPVLSYELFKYGGEGGIDLSRSCGPANHCVHGL
SNMISPVPEPSGLIPVCAIYEKARTFRALL"
complement(2111. .2198)
/gene="serx"
/note="synonym: b1032"
complement(2111. .2198)
/gene="serx"
/product="rRNA-Ser"
/note="anticodon: GGA; CG Site No. 17878"
/function="RNA; rRNA"
2269. .2236
/note="factor Sigma70; predicted +1 start at 1096980"
2338. .2365

/note="factor Sigma70; predicted +1 start at 1097049"
2355. .2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393. .3370
/gene="ycdu"
/note="synonym: b1033"
2393. .3370
/gene="ycdu"
/function="putative enzyme; Not classified"
/note="0325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPA_CORGL
SW: P45637"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="WRIINFAORFSGVSDIIFYPHTPDYOMWIBALRAIPQAVRAW
KSGNDSADVALVWHPVEMAGRDIAKAVFALGAGVDSILSKLQHPMLNDSVLPFR
LEDYMGEOHMOEYAVSQTLMHFRFPDRIOQNSHNPPEYHREDFTIGLGGVL
GSKVAQSLQIWRPFLRCVSRTRKSPGVQSFAGBELSAFLSQCVLITLNPETV
GIHQQLLEKLPGDAVYLTNLARGVHVEDDLAALDSQKVGAMLDVNERBPLPESP
LMQHPRTVITPVAALTRPAVAVEISRTIAQLBKGRVCGQVDRARGY"
3278. .3308
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1097992"
3383. .3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396. .3423

Query Match 99.6%; Score 454.4; DB 1; Length 10346;
Best Local Similarity 99.8%; Pred. No. 3.9e-121;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGAACCTTTAAAGTGAAGCAATTCGATTCCTCGGTAGCGCTTCGCA 60
8993 ATGAACCTTTAAAGTGAAGCAATTCGATTCCTCGGTAGCGCTTCGCA 9052

61 GGTGTTGTTCTCAATACGCGCGCGCGGTAAACCAAGGTGTCGCGTAATATAGCGGC 120
9053 GGTGTTGTTCTCAATACGCGCGCGCGGTAAACCAAGGTGTCGCGTAATATAGCGGC 9112

121 CCAATTCGACGGAACATTACAGAGTGGCGGTACTCTGCACTTGCTGGAA 180
9113 CCAATTCGACGGAACATTACAGAGTGGCGGTACTCTGCACTTGCTGGAA 9172

181 ACTGATGCCCGTAACCTGACTGATTAACCAAGATGGCGGCGTAATGTCAGAT 240
9173 ACTGATGCCCGTAACCTGACTGATTAACCAAGATGGCGGCGTAATGTCAGAT 9232

241 GTTGCTAGGGCTAGATGACAGTCAATGATTCGACCAACGTCCTCGGTAAACGC 300
9233 GTTGCTAGGGCTAGATGACAGTCAATGATTCGACCAACGTCCTCGGTAAACGC 9292

301 GCTACTCTGATGATGAGGAACGCAAAATCTGAATGAGGTAAACAGTTCGCTGT 360
9293 GCTACTCTGATGATGAGGAACGCAAAATCTGAATGAGGTAAACAGTTCGCTGT 9352

361 GGCACGCGTGTGAGTTCACCAAGTCACTTAATCTCCGTCAACGTGACTCAGTT 420
9353 GGCACGCGTGTGAGTTCACCAAGTCACTTAATCTCCGTCAACGTGACTCAGTT 9412

421 GGCCTTGGTAAACAAGCGACCGCTCATCAGTAA 456
9413 GGCCTTGGTAAACAAGCGACCGCTCATCAGTAA 9448

RESULT 5
D90741 15047 bp DNA linear BCT 25-DEC-2002
LOCUS Escherichia coli K12 genomic DNA. (23,7 - 24,0 min).
DEFINITION D90741 AB001340
ACCESSION D90741.1 GI:1651509
VERSION

KEYWORDS

Complete and shotgun sequencing; csqG; csqF; csqE; csqD; csqB;

SOURCE

csqG; ycdE; cIs; nov; mdog.

ORGANISM

Escherichia coli K12
Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE

1

Oshima, T., Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
Nishio, Y., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K.,
Wada, C., Yamamoto, Y., Yano, M., and Horiiuchi, T.,
A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiiuchi, T.,
Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.

TITLE

The systematic sequencing of the Escherichia coli genome in Japan

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

The Japan E.coli genome DNA sequencing project

Group: The Japan E.coli genome DNA sequencing project

Members: (1995.4 - 1996.3)
Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Horiiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,
Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.Headed by:
Name: Takashi Horiiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotsada Mori
Address: NARA Institute of Science and Technology,
Ikoma, 630-01, Japan
E-mail: hmori@nara.ac.jp
URL:
The Japan E. coli genome database
http://swj.aist-nara.ac.jp

FEATURES

SOURCE

Location/Qualifiers
1. 15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the
initiation site of phage (0 min.). This clone is from
Kohara lambda miniset library."

CDS

CDS

CDS

CDS

CDS

gene

CDS

<1. 276
/note="unnamed protein product; ORF_ID:0230#5"
/codon_start=1
/transl_table=11
/protein_id="BAJ35822.1"
/db_xref="GI:4062605"
/translation="DLFPSSNNALSTFPANPLRLFILVEMWITLVSLATLVE
WRIMEAFELICVPCILTYFFYCRFLMNDPMMACDMYFRMGHPK"
1152. 2090
/note="ORF_ID:0231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in prob-proA
intergenic region"
/protein_id="BAJ35823.1"
/db_xref="GI:4062606"
/translation="MDIIPYPTPTQWMEALRAKAIPOARVAMKSGDNDADYALV
WHPVEMLAGRLKAVFALGAGVDSILSKLOHPMTMPSVPLRIEDTQGEOMBY
AVSQVLAHWREREDYRIQONSQHWQPLREYREDPTGILGAGVLSKVAQSLQTFE
PLRCWERTKSWPGVQSPFAGRELSAFISCRVILNLENTPEVIGLIINOOLEKLPD
GAYLNLARGVHVEDDLAALDSKRYGAMLDVFNPNPEPESPLWQHPVITTPHY
AATTPAEALVEYISRTIQLKEKERVCGQVDRANGY"
2145. 2882
/note="unnamed protein product; ORF_ID:0231#2"
/codon_start=1
/transl_table=11
/protein_id="BAJ35824.1"
/db_xref="GI:4062607"
/translation="WYVDLHMHTVASTAVSTISDYIAQAKGKIKLPAITDHPDM
EDAPPHWFIIMRIIMPRVIVGVCGILREIENIKNVGSEIDSCMKPDSLLIAGHRE
PYFAPDKATNTQMTATISGNVHTISHGNKRYIDVAVVAEAAKQVLEINS
SELSHRSKSPNCEVNAVAVDAGWVALGSDSHTAFTWGEFEECKIADVDEPPER
ILNYSRRLNLFESRGMAPIAEFADL"
2906. 3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein H11543"
/protein_id="BAJ35825.1"
/db_xref="GI:4062608"
/translation="WNEFSILCRVLSLYRQDPLVPLFTLLIREKLAANPLEO
DELUTLQKSCMTQVSADYNALFTGDECAVPPRSANVETAPVRAFTISEGMP
ADTPADHIGITLLAASWLEDSDESEALLETLPSEVILPWCAGFLGVEAHATTPEW
RTVAPLRDAISAMWDELDEEDSE"
3514. 4053
/note="unnamed protein product; ORF_ID:0231#4"
/codon_start=1
/transl_table=11
/protein_id="BAJ35826.1"
/db_xref="GI:4062609"
/translation="MCSISFPAFAFAMRGNNILSIATITGILSGIKWVAVSLGL
SWAGFLGCTAYFACPOGILKGLAISATLISGVYMANVITVIGALAPHEILGVYTG
IVAFMCIQAKOLLSLFPVGTGACATPAGQGMKLVLSLALGLIFGVMKNSGLW
LAARSAKTAHREDEIKKKA"
complement(4117. 4950)
/gene="csqG"
complement(4117. 4950)
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsqG precursor"
/protein_id="BAJ35827.1"
/db_xref="GI:4062610"
/translation="MQRFLVAVWMLSCCLTAPPKEAAPTMPRAQSYDOLTHLPA
PTKRTFVSYNIDETGCFKPPASNFSPVAVQSSATLVYALAKDSRFITLEQGLQ
NLNERRKIRIAQENGVAINNRIPLGISTANIVBESITIGSNVKSQGVGARYG
IGADTOYDOJAVNIRVNVSTGRTSLSSVSTKITSITLSEVQAGVFPFIDQRLTGE
VGYSNPEPMLCLMSAIEIGVYFLINDIGLWLDLONKARQNDIIVKRYHMSVPEP

```

gene      S"
          complement(4977. .5393)
          /gene="csgB"
          /complement(4977. .5393)
CDS
          /gene="csgB"
          /note="ORF_ID:0231#6
          similar to SwissProt Accession Number P52104"
          /codon_start=1
          /transl_table=11
          /product="Curli production assembly/transport component
          csgB precursor."
          /protein_id="BAJ35828.1"
          /db_xref="GI:4062612"
          /translation="MRVKHAYVLMILISPLSMAGMTPEOPNPNGPNNGAPILNS
          AQONSXYKDPSPYNDPFGIETPSALDNLFTQALQSQILGGLLSNINTGKRGKRVNTDIYI
          DIANKDQQLAVLDKRTGTSTIQVSGLQNNSTDF"
gene
          complement(5418. .5807)
          /gene="csgE"
          /complement(5418. .5807)
CDS
          /gene="csgE"
          /note="ORF_ID:0231#7
          similar to SwissProt Accession Number P52105"
          /codon_start=1
          /transl_table=11
          /product="Curli production assembly/transport component
          csgE precursor."
          /protein_id="BAJ35829.1"
          /db_xref="GI:4062612"
          /translation="MKRYLRWIVAAEFLFAAGNLHAEVEVPGLLTDHTVSSIGHDFY
          RAFPDKMSDYDGLNLTINERSAWGSWITTVNDVIFQTFLEPKKDFPKTVVFAI
          IOTEBALNRQIOLNLSGLADHEP"
          complement(5812. .6462)
          /gene="csgD"
          /complement(5812. .6462)
gene
          /gene="csgD"
          /note="ORF_ID:0231#8
          similar to SwissProt Accession Number P52106"
          /codon_start=1
          /transl_table=11
          /product="Probable csgB operon transcriptional regulatory
          protein."
          /protein_id="BAJ35830.1"
          /db_xref="GI:4062613"
          /translation="MFNEVHSIHGHITLLITKSSLOATALLQHKQSALITKGLNLTQ
          RSLDDISSGSITLLDMEADKKLIHYMDTISRKNKIKILLTPEYPRIDENMP
          HINGVFYSMEDQERVVNGVGLGVEGECYFTQKLSYLLTHSGNRYNSTESALLTHRE
          KEINKLRIIGASNNELASLPISNTVKTHTLNLFKLIIVKRTQAVSWANDNLR"
          7217. .7672
          /gene="csgB"
          /gene="csgB"
          7217. .7672
CDS
          99.6%; Score 454.4; DB 1; Length 15047;
          Query Match
          Similarity 99.8%; Pred. No. 4.1e-12;
          Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Db 7953 GTTGTCAGGGCGTCAGATGACAGCTCAATGCATCTGCACCAACGNGCTTCGCTAACAC 8012
Qy 301 GCTACTTTGATCAGTGGAAACGCAAAATTTCTGAATGACGTTTAAACAGTTCCGTGCT 360
Db 8013 GCTACTTTGATCAGTGGAAACGCAAAATTTCTGAATGACGTTTAAACAGTTCCGTGCT 8072
Qy 361 GGCAACGGTGCCTGACGTTGACAGCTGACATCTAATCTCTCCGTCGAAGTGACTAGGTT 420
Db 8073 GGCAACGGTGCCTGACGTTGACAGCTGACATCTAATCTCTCCGTCGAAGTGACTAGGTT 8132
Qy 421 GGCTTTGGTAAACAACGACCGCTCATCATGACTTAA 456
Db 8133 GGCTTTGGTAAACAACGACCGCTCATCATGACTTAA 8168

RESULT 6
AF275733 1711 bp DNA linear BCT 15-MAY-2001
LOCUS AF275733
DEFINITION Escherichia coli str. 43895 Red Variant CsgB protein (csgB) and
ACCESSION AF275733
VERSION AF275733
KEYWORDS csgB, complete cds.
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 1711)
Ulrich, G.A., Keen, J.E. and Elder, R.O.
Mutations in the csgD promoter associated with variations in curli
expression in certain strains of Escherichia coli O157:H7
Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
JOURNAL MEDLINE
21218556
PUBMED 11319125
REFERENCES
2 (bases 1 to 1711)
Ulrich, G.A., Keen, J.E. and Elder, R.O.
Direct Submission
JOURNAL Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES
source
1. 1711
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
757. 1212
/gene="csgB"
757. 1212
/gene="csgB"
757. 1212
/gene="csgB"
/codon_start=1
/transl_table=11
/product="CsgB protein"
/protein_id="AAK53211.1"
/db_xref="GI:14039400"
/translation="MKRLKPMILITIGARCIATAAAGYDLSNRYNPAVNELSSRPN
OALITGGAGTNSAQLRQGSKLAIVAQESSRRAKIDGTGDLNLAVITDQASANDA
SISQAGANTAMIIQKSGNKNANTITQGTAKTALVOROSOMAIRVOR"
1253. 1711
/gene="csgA"
1253. 1711
/gene="csgA"
/codon_start=1
/transl_table=11
/product="CsgA protein"
/protein_id="AAK53212.1"
/db_xref="GI:14039401"
/translation="MKRLKVAATAAIVPSSALAGVVPYQGGGNGHGGGNGSGNPS
ELNTYQGGGNSALALQADARNSDLTITQGGGAGAVGGGSDSSITDLTQGGNSA
TLQDWNKDSHMTVKQFGGNGAAGVDDTANSTVNTVQVGNNAATAAQY"
ORIGIN
Query Match 86.6%; Score 395; DB 1; Length 1711;

```

QY	1	ATGAAACTTTTAAAGTAGAAGCAATTGACGACATGCTATTTCTCCGTAGCGCTTGGCA	60
Db	1253	ATGAAACTTTTAAAGTAGAAGCAATTGACGACATGCTATTTCTCCGTAGCGCTTGGCA	1312
QY	61	GCTGTTGTTCCGAGTACGGGGC--GGCGGTACACCGGTGGTGGCGGTATATATAGC	117
Db	1313	GCTGTTGTTCCGAGTACGGGGC--GGCGGTACACCGGTGGTGGCGGTATATATAGC	1372
QY	118	GGCCCAATTTCTGAGCTGAACATTATCCAGTACGGGTGGGTAACTTGTGCACTTGCCTG	177
Db	1373	GGCCCAATTTCTGAGCTGAACATTATCCAGTACGGGTGGGTAACTTGTGCACTTGCCTG	1432
QY	178	CAAACTGATGCCCGTAACCTCTGACTTATTTACCGACATGGCGCGGTAAATGGTGCA	237
Db	1433	CAAACTGATGGCTGTAACTCTATCTATCTATTTACCGACATGGGTGGTGGTAAACGGTGCA	1492
QY	238	GATGTTGGTCAGGGCTCMAGTACACGCTCAATCGATCGAACCCAAAGCTTGGCTGTAAC	297
Db	1493	GATGTTGGTCAGGGCTCMAGTACACGCTCAATCGATCGAACCCAAAGCTTGGCTGTAAC	1552
QY	298	AGCGCTACTTTGATCACTGATGCAATGCAACGGGCAAAATTCTGAAATGACGGTTAAACAGTTCCGT	357
Db	1553	AGCGCTACTTTGATCACTGATGCAATGCAACGGGCAAAATTCTGAAATGACGGTTAAACAGTTCCGT	1612
QY	358	GGTGGCAACGGTGTGTGAGTTGACCAAGACTGCATCTAACTCTCCGTCAACGTGACTGAG	417
Db	1613	GGTGGCAACGGTGTGTGAGTTGACCAAGACTGCATCTAACTCTCCGTCAACGTGACTGAG	1672
QY	418	GTTGGGCTTGGTAAACAACGGACCGGCTCACTAGTACTTAA	456
Db	1673	GTTGGCTTGGTAAACAACGGACCGGCTCACTAGTACTTAA	1711

```

/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
617..1072
/gene="cesg8"
/note="synonym: Z1675"
617..1072
/gene="cesg8"
/function="structural component; Cell exterior constituents"
/note="Residues 1 to 151 of 151 are 100.00 pct identical to residues 1 to 151 of 151 from Escherichia coli K-12 Strain MG1655; B1041"
/codon_start=1
/transl_table=11
/product="minor curlin subunit precursor, similar to CsgA protein, id=AA655787.1"
/db_xref="GI:12514573"
/translations="MKMKLLFLMVLITIGAPGIAAAGYDIDANSEYFAVNEISKSEFSEFIAAIGAGCGNNSAQRGSGSKLFLAAVAGSSNRAKIDDTGYNLAITIDGASNDALISGAGVMTATIKSGSNANITGYGTAKIAIVGRSQAIRVQR"
113..1571
/gene="cesgA"
/note="synonym: Z1676"
113..1571
/gene="cesgA"
/function="regulator; Cell exterior constituents"
/note="Residues 1 to 152 of 152 are 96.71 pct identical to residues 1 to 151 of 151 from Escherichia coli K-12 Strain B1042"

```

[illegible]

```
/mol_type="genomic DNA"
```

```

cryptic
/protein_id="AAG55788.1"
/db_xref="GI:12514574"
/translation="MTLLKVAIAIAIVFSGSLAGVYPOVGGGGGGGGGCGGNNSPGNS
ELNIWKKYGGGNSLALQADARNSDLTIQHGSGNADVGQSDSDSLTQCGFGNS
TLDWNGKSGSHVTVKQFGGNGAAVDQTSNSTVWTVQVGGNNATAHQY"
160..1962
/gene="csqC"
/note="synonym: Z1677"
1630..1962
/gene="csqC"
/function="putative structure; Not classified"
/note="Residues 1 to 110 of 110 are 98.18 pct identical to
M01655: B1043"
/codon_start=1
/transl_table=1
/product="putative curJ production protein"
/protein_id="AAG55789.1"
/db_xref="GI:12514575"
/translation="MNALLALSSQITFNTTQGDWYTIPEVLTNQCCLRQVIL
SLRGGSGSQGTQKQETLSLPANQPIALTKLISNPDRVKIVTVSDGQSLHSQ
2093..2394
/gene="Z1678"
2083..2394
/gene="Z1678"
/function="orf: Unknown function"
/note="Residues 1 to 103 of 103 are 98.05 pct identical to
M01655: B1044"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAG55790.1"
/db_xref="GI:12514576"
/translation="MPPPIFNSIMLGSLEFPPIAIGTAGTGAGGVTHFYQIYPAQDV
STGSPVEMNCPQWGSVGGKTVSSKALMSGVYNAQIIVAKYQIDKQKGAVMNIEY

```


TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 derived from the Sakai outbreak
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE	20198780
PUBMED	10734605
REFERENCE	2
AUTHORS	Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.
TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic <i>Escherichia coli</i> O157:H7 Sakai strain and an <i>Escherichia coli</i> K-12 strain MG1655
JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE	20557356
PUBMED	11108008
REFERENCE	3
AUTHORS	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasekawa, C. and Shinagawa, H.
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 strain derived from the Sakai outbreak
JOURNAL	Gene 258 (1-2), 127-139 (2000)
MEDLINE	20564182
PUBMED	111101050
REFERENCE	4
AUTHORS	Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasekawa, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
TITLE	Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL	DNA Res. 8 (1), 11-22 (2001)
MEDLINE	21156231
PUBMED	11258796
REFERENCE	5
AUTHORS	Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp), Tel: 81-6-6879-8365, URL: http://www.gen-info.osaka-u.ac.jp/, Fax: 81-6-6879-2047)
COMMENT	genome project.
FEATURES	location/Qualifiers
source	1..327773
	/organism="Escherichia coli O157:H7"
	/mol_type="Genomic DNA"
	/strain="O157:H7"
	/sub_strain="RIMD 0500952"
	/db_xref="taxon:83334"
gene	150..2258
	/gene="ECs1032"
	/note="similar to YCBY ECOLI g11787181 percent identity 99 in 702 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="putative oxidoreductase"
	/protein_id="BAB34455.1"
	/db_xref="GI:13360492"
	/translation="MNSLPASTRAGLEELIKTELENLAIVECQVQVGVHFKDTRIVYGLMWSRLASRLMLPLGRCVYSDDLVLGVQAINWEMFPGATFVHFGSLNDITV RNSQYAMKVDALVDAFTRKMLPENNVDRAFDIRVWVWLHKEFASIALDSDGHL LKRIYRAGIADIKETLAAILVMSQWPGTPLDPMCGSGTLLIIEAALDADPAGL HGRWFGSCMAQIDEALIQEVAEAKAETARKKGLAYSSHPFGSGSDARVIOPRATNKL GIGISLITFEVADVAQLTNPPLPKBYGTIVLSNPYGRLLDSEPLALIALSLGRIMK
CDS	
gene	270..4177
	/gene="ECs1033"
	/note="similar to UTP ECOLI g11787182 percent identity 99 in 635 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="putative ATP-binding component of a transport system"
	/protein_id="BAB34456.1"
	/db_xref="GI:13360493"
	/translation="MSLSIMGAMLSFSDAPLIDNAELIHEDNERVICVGNAGKST IMKILIRREGLDGRIITYQDILVARLOQDPPRNVEGVYFVAGEIGVGNAGKRYH DLSRLVNDPSEKNINELAKVQEOLDHNLMOENRINEVIAQGLDNNVSLKSG WLRKALGRALVSNPRLIDEPNHLIDETIDLEGLPTNGITIFISDRSIFRN MATRIVDLRGKLVTPGVYDQYLKEXEALRVEELONAEPRDIAQOEVRIGRTA RRTNREGVRLKAMRERGERREVMGTAKOVBASRSGLIVBMEVCYQVQKGL VDFSAQVLRGDKIALIGPNCGKTYLLKMLGLOLDSGNHVGTELVYFQDRA ELDPKTVNDNLABKQVYMNVRKRRHVLGLOLDFPKPKAMPVRLASGSENRNL IARLPKSNLILIDEPNDLVEFTELEELIDSYQCTVLPKAMPVRLASGSENRNL ITEGKSGRIVGVGHDRAGQOEDVADLQKAVKNEPAPKAEYKRSSTKSYTL QNELPOLLEDEAKLEALQTVADASFPSCHEQKVLADMAAABEQLBQAFER MBYLALNKG"
gene	4307..5560
	/gene="ECs1034"
	/note="similar to POIA ECOLI g11787183 percent identity 99 in 417 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="paraquat-inducible protein A"
	/protein_id="BAB34457.1"
	/db_xref="GI:13360494"
	/translation="WCEHHHAHHLICQCDMLVALPRLEHGKACPRCGTTLTVAM DAPRQPTAYALALFMLLSNLPFNMMVAGVSEITLLEIRGVLFSEDAVSLGTF FLFVQVLPARFCITILLVNRALPVRLEKOLARVLPOLKTMGAETIACVIVSV KLAAYGSIQVSSFLPWCIFCVQLAFQCVDRMLMDIADPELRQPKGVGIR QGLRSQSCCTAILPADPVCPCGCTGVRRRSLQMTLALVTSIMLTPANILIPM VTDLGSKMSTIILAGVITILMSRGSYPVAVIFLSIMVPTLMIAVLCPWDAGHG KRDSERHMLTYVEVERPGRKSNIDVFIVLVSALVMGGLMSITYPAMGLMALVIVM TMSAMTFDRRLWDRQSEHES"
gene	5565..7205
	/gene="ECs1035"
	/note="similar to POIB ECOLI g11787184 percent identity 99 in 546 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="paraquat-inducible protein B"
	/protein_id="BAB34458.1"
	/db_xref="GI:13360495"
	/translation="MESNNGRAKIOKYNMSPWIFPVTALICAWLFYHSHQGBE VVWVLPQGRGSGIGLSTLGSVYIELQPAKSKRMKXDLSDPLAPDAGIRV VDSKKAQGLSGDVLFRGVSVESITDQKRNISYQLFINAPYDRLVTSNVP WDSIAYHDLISAGRVNMGSLITLLSGVGFVDEGLDQGPAPATAPLVDXOS IODSIYATDILPAGKDSVGLQGAVERRGIRLGVSVVPPFAPNMOTPRDDYR IPLVIRIPEERIKMQLGNADVHHLGELIKRGISLKTGNLTATGALYYDLYFVNT PATIGIRENGVQIIPYVSGGLAQICQRLMALDKINKPLPMPNIEGATSTLESQRT MNKQITLDSNKLILASGMOQLPDMQSTLREINSMQGFQPSGAAYNNVADMDQRL
CDS	

Query	Match	Best Local Similarity	86.6%; Score 395; DB 1; Length 327773;
Matches	426; Conservative	0; Mismatches 30; Indels 3; Gaps 1;	
QY	1	ATGAACCTTTTAAAGTAGAAGCAATTCGACGCAATGTAATTCCTCCGCTAGACGCTCTGGCA	60
Db	326421	ATGAACCTTTTAAAGTAGAAGCAATTCGACGCAATGTAATTCCTCCGCTAGACGCTCTGGCA	326480
QY	61	GATGTTGTTCCCTCAGTACGGCGGCG--GGCGGTAACCAAGTGGTGGCGGTAATTAADAC	117
Db	326481	GATGTTGTTCCCTCAGTACGGCGGCGGCGGTAATTAADAC	326540
QY	118	GGCCCAATTCGAGCTGAACATTTACAGTACGGTGGCGGTACTCTGCACCTTGGCTTG	177
Db	326541	GGCCCAATTCGAGCTGAACATTTACAGTACGGTGGCGGTACTCTGCACCTTGGCTTG	326600
QY	178	CAAACTGATGCCCGGTAACCTGCACTGTAATTAATCCACACATGAGCGCGGTAATGGTGA	237
Db	326601	CAAACTGATGCCCGGTAACCTGCACTGTAATTAATCCACACATGAGCGCGGTAATGGTGA	326660
QY	238	GATGTTGGTCAAGGGCTCAGATGACAGCTCAATCGATCTGACCAAGTGGCTTCGGTAAAC	297
Db	326661	GATGTTGGTCAAGGGCTCAGATGACAGCTCAATCGATCTGACCAAGTGGCTTCGGTAAAC	326720
QY	298	AGCGCTACTCTTGATCAGTGAACCGGCAAAAATTCGAATGAAGCGGTTAAACAGTTCCGT	357
Db	326721	AGCGCTACTCTTGATCAGTGAACCGGTAAGCTCATATGACAGTGAATTAACAGTTCCGT	326780
QY	358	GGTGGCAACGGGTCGACAGTTGACACAGATGATCAATCCCTCCGCTCAACGGAATCGAC	417
Db	326781	GGTGGCAACGGGTCGACAGTTGACACAGATGATCAATCCCTCCGCTCAACGGAATCGAC	326840

Cyt 418 GTTGGCTTTGGTACACGAGCCGCTATCAGTACTAA 456
 |||||||
 Db 326841 GTTGGCTTTGGTACACGAGCCGCTATCAGTACTAA 326879
 |||||||
 RESULT 9
 AEO16759 306358 bp DNA linear BCT 24-DEC-2002
 AEO16759
 Locus
 Definition Escherichia coli CFT073 section 5 of 18 of the complete genome.
 Accession AEO16759 AEO14075
 Version
 Keywords
 SOURCE
 ORGANISM
 Escherichia coli CFT073
 Escherichia coli CFT073
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE
 1 (bases 1 to 306358)
 Authors Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
 Raeko,D.A., Bucklee,E.L., Liu,S.-R., Boutin,A., Hackett,J.,
 Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
 Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
 Extensive Mosaic Structure Revealed by the Complete Genome Sequence
 of Uropathogenic Escherichia coli
 Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
 TITLE
 JOURNAL
 PUBMED
 12472157
 2 (bases 1 to 306358)
 Authors Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
 Raeko,D.A., Bucklee,E.L., Liu,S.-R., Boutin,A., Hackett,J.,
 Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
 Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
 Direct Submission
 Submitted (20-JUN-2002) Genetics Laboratory, University of
 Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
 FEATURES
 source
 location/Qualifiers
 1. 306358
 /organism="Escherichia coli CFT073"
 /mol_type="genomic DNA"
 /strain="CFT073"
 /db_xref="taxon:199310"
 complement (83. 3820)
 /gene="iroC"
 /locus_tag="C1253"
 complement (83. 3820)
 /gene="iroC"
 /locus_tag="C1253"
 complement (83. 3820)
 /gene="iroC"
 /locus_tag="C1253"
 /note="Residues 27 to 1231 of 1245 are 80.16 pct identical
 to residues 1 to 1209 of 1218 from GenPept.129 :
 >emb|CADO5883.1| (A1627276) putative ABC transporter
 protein [Salmonella enterica subsp. enterica serovar
 Typh1]"
 /codon_start=1
 /transl_table=11
 /product="ATP binding cassette (ABC) transporter homolog"
 /protein_id="AAV79727.1"
 /db_xref="GI:26107528"
 /translation="MRDRSLAFLRNDPASGSYFTDGGILMPANHPTTPAOSWIVRIA
 RYMERKLSIVIVVASWSTYIIAALTPLTRQAVNDALAGNPARIPMLACGLILAF
 FDTIGNVRRGVAGMLSLMWQHTLRGRVTSIOKDGAGDQDARTGVSRTSDIQ
 VHTLLQMCPLAIVPTVYIAGTAVVMSMPMTLVYCTVCLATLAARPRVQOT
 GMSDQALNTEHIREVLAQIIYVSKSVAMRETHLIDROSRIYVRIGAVTSQAP
 GAATMLAPVIGQIVLLCYGQWSVHGRIDLGTFVAASPLAMLTGTRVLASVLAQ
 RTQASVERFALIDTDSOMEDGTESINSQVGLLENMFSVDYHGRDLISDISFLR
 AGTAVVAVGSGSGSTLLMLLARPYDPCSGKWLMTSGRQRLDRIEALARRVGI
 VEGDAIFAGTAVENIAYGHPQATDDIDRIAAAGASFINALPGKPSQILTERTN
 LSGGQQRIALAALITAPDVLIDDTTDSANDAVYEAINTALAGRYADSGHMLIVLAR
 RSTTLQASRVVLDGKRWDTGTPABLEKRCFAFLMTGSDFLATSHNSHELP
 AEAATDDVDVTDGDKFVAKTRVPENAVQALAGKRVTSILKPEVAMPFYAALI
 ALDSAGVGVLLDQIGIDSGVAGDMSTIGLCALALCLVIYVGCSTYLLQTFARA
 AEAIVQSVLRISFGHMLRLGLPMHEKQADSRILRMTVDVSLARFLONGIAGAATSLV
 TMAVLAATMFWLMDPIALATLALGASVPVAALATMYRLSTPVAQAELTIGKNSTQES
 KSGGLAVVQSHQGELEGARLRLSRPRATPRACKRYLAIVYPTPTCTEASVAVL
 LVASAGVAGEKMTAGVLAIFLLGLGQYVQVQDSIVQADQVAVSGKHIDBLATE
 GTNLSSSSVLPVYGLALHDEYFSPDSEHSHINLTLTLTPEGMAVAVAGRGSAKS

TLTKLGIPTPTGHNIRIGVOMIDNASLIERYRQIGLVQDVALLPSSDIAENIRYR
PSATNEVEIISQAGLIEVNCILPQGFRTPVNNGADLSAGOROLITAPOLANAH
ILLDEATISCLDRISERLMSLTDVHAKHSALIVAHRLTTAQRCDLITAVIDKLL
AEYGHQQLSAGGLYTRLMWDSVSTALHRQNMKEETFG"
/complement (3882..5045)
/gene="iroB"
/locus_tag="c1254"
complement (3882..5045)
/gene="iroB"
/locus_tag="c1254"
/note="Escherichia coli O157:H7 ortholog: z1190"
/locus_start=1
/transl_table=11
/product="Putative glucosyltransferase"
/protein_id="AAN79728.1"
/db_xref="GI:26107529"
/translation="MRRLEPDLRQAEERDPLMRILFVGPPLYGLVPLVSLAQAFVNG
HEVLASGGQFAQKAAGLVVDAPGLDSAGYRHHQKRSNIGTOMNPFPS
EMADHLVEFAGHMRPDLITPLVIGLITAKYDIPVMQTVRGHTRPHIRGVT
SLTDAYRRHNVGATPRDMAMIDVTPSMSTLNDGPRIIPQYVYNGGAVWEWVER
RPDRKRLVSLGTVKPMVDGLIANNMWSASEVDREIILHISANARSDLSLSENR
LVDMIMGVPLNADGFIHGGAGNTLTLAHGIPQIVFGQADPRVAVVAERCG
IIPGVGLSSNMIMAFNNRSLRKASEEVAEMAAPCEVAKSLITVQKG"
5535..5834
/locus_tag="c1255"
5535..5834
/locus_tag="c1255"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79729.1"
/db_xref="GI:26107530"
/translation="MKPTLQTVENICQILFVLEPLNGTLIDVEALVSCSVAKNQ
RPPASVSSSVYASVSRADIMLSNNPLFAPQVSPFCGSLAAGSHVITLE"
5867..6577
/insertion_seq="IS1541A-like"
5961..6470
/locus_tag="c1256"
5961..6470
/locus_tag="c1256"
/note="Residues 18 to 169 of 169 are 94.07 pct identical
to residues 1 to 152 of 152 from Swi58Prot.40 :
>sp|O57334|T200_SALT Transposase for insertion sequence
element IS200"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79730.1"
/db_xref="GI:26107531"
/translation="MRSGNCKCQTRNCKGVPMNGESLAHTRNCKYHIVFAKYRQ
VFRERKRAICILKLCWMSVRIIEACCADHIMLVEIPKMSVSGMGYLGKS
SLMPEQGDIKFYRNRFEWRCGYVDVAGKNTAKIDYIKHQLEEDKMGQELISIPY
PGSPFTGRK"
complement (6578..6940)
/note="incomplete"
/insertion_seq="IS4"
complement (6598..6960)
/locus_tag="c1257"
complement (6598..6960)
/locus_tag="c1257"
/function="IS, phage, Tn"
/note="Residues 8 to 120 of 120 are 98.23 pct identical to
residues 316 to 428 of 442 from M6165 : b4278"
/codon_start=1
/transl_table=11
/product="Putative conserved protein"
/protein_id="AAN79731.1"
/db_xref="GI:26107532"
/translation="MPROPDGGYREIKOTMOSRLTRSKKPELYEQSLMGVLAIVNL
VRYQIMAEHLKGVMPNOLSPSCGMWRMLTLQGASPRITELMRDLASMGQLV
KLPRRETAPEPRVYKERP"
7651..8013
/locus_tag="c1258"

CDS
7651..8013
/locus_tag="c1258"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79732.1"
/db_xref="GI:26107533"
/translation="MGACFYSWGTSEFYGFPMWGAELVIMMPEACFMILQSLYS
LDGQVLFSTWELNKQYEGSEFYQIFSGNIVASCSGCVSSNSAHYARCTDFWRH
YSLNNVACQLGQACLS"
7835..8269
/locus_tag="c1259"
7835..8269
/locus_tag="c1259"
/note="Residues 1 to 139 of 144 are 50.71 pct identical to
residues 104 to 242 of 258 from GenPept.129 :
>gi|AK18487.1|AF348706_176 (AF348706)orf, hypothetical
[Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79733.1"
/db_xref="GI:26107534"
/translation="MKDSLFRPLALILVPPALEPPIRHITPGALITCVTAL
ITLLASLGRVSVKISLSEIFRRHRECMAGAFIMTFSLIAGWYRQFKEY
QTLHYENARYVILARYDGRVLVSGSYRSGSRFVINGHRL"
8477..8553
/locus_tag="c5534"
8477..8553
/locus_tag="c5534"
/locus_tag="c5534"
/product="tRNA-Arg"
/function="tRNA; tRNA"
complement (8832..9290)
/locus_tag="c1260"
complement (8832..9290)
/locus_tag="c1260"
/function="putative enzyme; Transposases Insertion
Sequence Associated"
/note="Escherichia coli O157:H7 ortholog: z4324"
/codon_start=1
/transl_table=11
/product="Putative Transposase"
/protein_id="AAN79734.1"
/db_xref="GI:26107535"
/translation="WASADARRIVSHYERMLIEYHKAMKSGGTCVSELRMQTRDN
LERNVVIAFAVIMLGIROGISSEIQRNDSCKKITPTFMKLMVYLEKQQLPSQRP
TLKVAULTGRWDSKRTGRGVMVMDGMFRIDQWVEGYPVWKSIDQET"
complement (9302..9667)
/locus_tag="c1261"
complement (9302..9667)
/locus_tag="c1261"
/note="Residues 1 to 35 of 121 are 97.14 pct identical to
residues 59 to 93 of 316 from GenPept.129 :
>gi|AK18437.1|AF348706_126 (AF348706) putative
transposase [Shigella flexneri]"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79735.1"
/db_xref="GI:26107536"
/translation="MAERIGEIQKRVITVCDREADIWHLHYKSHGQCFVYRTAONS
Query Match 83.84; Score 382.2; DB 1; Length 306358;
Best Local Similarity 91.1%; Pred. No. 7.1e-100;
Matches 418; Conservative 0; Mismatches 38; Indels 3; Gaps 1;
OY 1 ATGAACCTTTAAAGTAGAAGCAATTCGACATGCTATTCGCGGTAGCGGTGCGGA 60
Db 45269 ATGAACCTTTAAAGTAGAAGCAATTCGACATGCTATTCGCGGTAGCGGTGCGGA 45328
OY 61 GGTGTTCTCCTCAGTAGCGCGCGCGGT---AACACGGTGTGTGCGGTATATATAC 117
Db 45329 GGTGTTCTCCTCAGTAGCGCGCGCGGTGCGCAACACGAGTGTGTGCGGTATATATAC 45388

assembly-transport component, 2nd curli operon gi: 1787274
(278 aa). BLAST with identity of 97% in 257 aa. This CDS
contains an in-frame stop codon. The sequence has been
checked and is believed to be correct."

gene
CDS

```
/codon_start=1
/transl_table=1
/complement(4454..5662)
/locus_tag="SF1032"
/complement(4454..5662)
/locus_tag="SF1032"
/notes="Residues 1 to 402 of a 402 aa protein from Plasmid R100
ref: NP_052925.1"
/codon_start=1
/transl_table=1
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCELDIHDLSLYOCPELHAKRLNSLTLAGHALDCKLTTEL
GNLPPTKARTKNIKRIDLLGNRLHKLRLAVTRHASFICGNTMPVLVWMDIR
EOKRLMVLRSVLAHGRSVTYEKAFLPUSQCSKKAADPLADIASILPSNTPLIVS
DAGFVPMYKSVKSGMTWMSRVKQVADLGAEMKPIASIHDSWSSKTLGYKR
LTKSNPIKSCUILLIKSRKQKQSTRICHPSPIYASAKKEPMIATLPURIR
TPKQVNIYSKMOIEETFDLKSRLPAYGLRHSRTSSSRPIMLLIAMIQICML
AGVHAQKQGMKDHQANTVRNRNVLSTVRLGMEVLHSGVTTIRSDSLVAATLTQNL
FTHGVLGKL"
```

gene

CDS

```
complement(5995..6384)
/gene="csge"
/locus_tag="SF1033"
/complement(5995..6384)
/gene="csge"
/locus_tag="SF1033"
/notes="Residues 1 to 129 of 129 are 98 pct identical to
residues 1 to 129 of a 129 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287173.1"
/codon_start=1
/transl_table=1
/product="curli production assembly-transport component,
2nd curli operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRYLRWVAALFFFAAGNLHAEVEVGLTTHPTVSSIGHDYF
RAVSDWESDYGNLTINRPSRSGMTITTVQDVIFQTFLEPKRDEKTVFVAL
IOTEBALNRQQLALSTDLADHDF"
/complement(6389..7069)
/gene="csge"
/locus_tag="SF1034"
/complement(6389..7069)
```

gene

CDS

```
/gene="csge"
/locus_tag="SF1034"
/notes="Residues 11 to 226 of 226 are 100 pct identical to
residues 1 to 216 of a 216 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=1
/product="putative 2-component transcriptional regulator
for 2nd curli operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="MRSIKKMGFIMFENEVSHGHITLLITKPSIQATALLQHKOSL
AITGKLNIRSLDSSGSIILLDMKDKLTHWQDTLSKNNKNIKILLNTPD
YPRADINWPHINGVYAMDEORVYNGQVLRGEQYFQKLASVYLTHSGNRYNS
TESALLTHREKEITLNRIGASNREIARSLFISENTVTHYNLFKIAVNRTOAVS
WANDNLRR"
```

gene

CDS

```
/gene="csge"
/locus_tag="SF1035"
/766..8248
/gene="csge"
/locus_tag="SF1035"
/notes="Residues 10 to 160 of 160 are 99 pct identical to
residues 1 to 151 of a 151 aa protein from Escherichia
```

Query Match
Best Local Similarity 98.0%; Score 340.8; DB 1; Length 10370;
Matches 345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```
colli O157:H7 EDL933 ref: NP_287175.1"
/codon_start=1
/transl_table=1
/product="minor curliin subunit precursor, similar to CsgA"
/protein_id="AAN42658.1"
/db_xref="GI:24051320"
/translation="MYDQVGDNNKKNKLFWMLTITGAPGIAAAGYDLANSEYNPAY
NELSSPNQARITIGAGNTNNSAQILROGSGKLAIVAAQSGSSNRKAITDGTGYNLAYI
DQASNDASTISQAYGNTAMIIIOKSGMKANITGYGTQKTVVVOGRSGMAIRVTOE
```

Qy	105	CGTAAATATAGCGGCCCAATTCGTAGCTGAAACATTTACAGTACGCTGCGGTACTC	164
Db	9538	CAGTAAATATAGCGGCCCAATTCGTAGCTGAAACATTTACAGTACGCTGCGGTACTC	9597
Qy	165	TGCACCTTGCTTGCACAACTGATGCGCCGTACTCTGACTTATCCCGCATGCGG	224
Db	9598	TGCACCTTGCTTGCACAACTGATGCGCCGTACTCTGACTTATCCCGCATGCGG	9657
Qy	225	CGTAAATGAGGACAGATCTTGTCAGGGCTCAGATGACAGCTCAATGCATGACCCAACG	284
Db	9658	TGCTATATGCTGACATGTTGTGACGGCTCAGATGACAGCTCAATGCATGACCCAACG	9717
Qy	285	TGCTTCCGTGAACACCGCTACTCTTGATCAGTGAACGGCAAAATCTGAATGACGGT	344
Db	9718	TGCTTCCGTGAACACCGCTACTCTTGATCAGTGAACGGCAAAATCTGAATGACGGT	9777
Qy	345	TAAACAGTTCGGTGTGTCAGACGGTCTGCGTACCACTGACATTAATCTCTCCGT	404
Db	9778	TAAACAGTTCGGTGTGTCAGACGGTCTGCGTACCACTGACATTAATCTCTCCGT	9837
Qy	405	CACGTGACTCAGGTTGCTTGTGTAACAACGCGACCGCTCATCATCTACTAA	456
Db	9838	CACGTGACTCAGGTTGCTTGTGTAACAACGCGACCGCTCATCATCTACTAA	9889

RESULT 11
AE016981
LOCUS
DEFINITION
Shigella flexneri 2a str. 2457T section 4 of 16 of the complete
genome.
ACCESSION
AE016981 AE014073
VERSION
AE016981.1 GI:30040616
KEYWORDS
SOURCE
ORGANISM
Shigella flexneri 2a str. 2457T
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE
1 (bases 1 to 292504)
Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,
Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,
Schwartz, D.C. and Blattner, F.R. Complete Genome Sequence and
Comparative Genomics of Shigella
flexneri Serotype 2a Strain 2457T
Infect Immun. 71 (5), 2775-2786 (2003)
2 (bases 1 to 292504)
Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, W.,
Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A.,
Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S.,
Schwartz, D.C. and Blattner, F.R. Direct Submission
Submitted (13-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
LOCATION/Qualifiers
1..292504
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"

```

gene
/db_xref="taxon:198215"
/complement(223..441)
/gene="infA"
/locus_tag="S0884"
/complement(223..441)
/gene="infA"
/locus_tag="S0884"
/function="factor; Proteins - translation and modification"
/notes="residues 1 to 72 of 72 are 100.00 pct identical to residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
/codon_start=1
/transl_table=11
/product="protein chain initiation factor IF-1"
/protein_id="AAP16348.1"
/db_xref="GI:30040617"
/translation="MAKEDNIEMOGVLETLPTNTWFRVLENGHVTAHISGKMKRY
RILTGDRVTELPYDLSKGRIVRSR"
/complement(726..1430)
/gene="aaf"
/locus_tag="S0885"
/complement(726..1430)
/gene="aaf"
/locus_tag="S0885"
/function="enzyme; Aminoacyl tRNA synthetases, tRNA modification"
/notes="residues 1 to 234 of 234 are 100.00 pct identical to residues 1 to 234 of 234 from Escherichia coli K-12 : B0885"
/codon_start=1
/transl_table=11
/product="leucyl, phenylalanyl-tRNA-protein transferase"
/protein_id="AAP16349.1"
/db_xref="GI:30040618"
/translation="MRVLQSRHSIAPSPPEGALKEPGLALGDSLSPARLMAVOR
GIPWESGDDILMWSPPRAVLNPESELIHSKMRPKRSPRYTMVAAGOVIEGC
ASDRKCTIRGVGEAYHRLHLEGHASISIVKREDELVCGMYGAOGLTCGSMPS
RMENASKTALVFCBERFIGHGSKLIDCVLNDHTASIGACEIPRDYLYINLQMKLGR
LPNNFWPRCLFSPQE"
/complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
/complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
/function="transport; Transport of small molecules: Other"
/notes="residues 1 to 573 of 573 are 97.03 pct identical to residues 1 to 573 of 573 from Escherichia coli K-12 : B0886"
/codon_start=1
/transl_table=11
/product="ATP-binding component of cytochrome-related transport"
/protein_id="AAP16350.1"
/db_xref="GI:30040619"
/translation="MRALPPLALYKHKMUSIGIVLAITYTLASIGLTLSCWPLS
LSAAGVAGLISFNVMPLPAAGVGAATRTAGRFERLVSDATFRVQHLRIYFSK
LPLSPAGIARQSEELNRVAVDDTLDHLYLRVISPVAFAFVIMVYGLSFLDF
TLATLGSVMLTLFLMPPLFYRAGSKTGONLTHLRGOOLAMLOGAELTFGA
SDRYTOENTEIOMLEAORROSELTALSGOIMLLGALAVILMMASGCVGNAOP
GALITAPVFCALAPREALAPYTGAFQHLGOVITASAVRSIDTDKRETPPDOTRVA
DRVSTLRDVOPTTPEOSQALKGISLVNAGEHAIIGRTGCGSTLQULTRAMPD
QQGEILLNDSPIASINENALROTISVVPQARFSAITLRLNLLASPOSSDBALSEL
RVRGIEKLBDAGLNSWLEGGRQLSGELRLAIARTLHDAPLVLDDETEGIDAT
TESQIELLAKMKREKTVLMTWHLRIGLSRQOIIVMDNGQIIEGTHAELLARQRY
YQPKQGL"
/complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"
/complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"

```

```

gene
/function="transport; Not classified"
/notes="residues 1 to 588 of 588 are 92.85 pct identical to residues 1 to 588 of 588 from Escherichia coli K-12 : B0887"
/codon_start=1
/transl_table=11
/product="ATP-binding component of cytochrome-related transport"
/protein_id="AAP16351.1"
/db_xref="GI:30040620"
/translation="MNKSRQKELTRMLKQOSVISORMLNISRLGFPVSGIILIAQWF
MARLIQHMIMENIPREALLPETLVTFVLAWVMLREBYGACOHIFRAIRYQ
LDRLOQGPAMIOGKPPAGSMATLVBODIDHDVYARLPOMALAVSPLILVAVFP
SNMAAILIIGTAPLILPLFPAUVGMAADNRNRFALARSGFIDRLRMETIRIF
GRGAETESTRSASEDFRQRTWEVRLAFLFSGGLEPFTSLIALVAYFGPSYUGEL
DFGHYDGVTLAAGFLALILAPFFQPLRDGTGYHAKAQAAGADSILKTMETPLAH
PQGEAEIALDPLTILBAEDLFTSPSGKTLAGLNTPLPAGQPAVLVGRSGGSSL
LNAISGFLSYGSLRINGIELRDLSPSMRKHLSWVGNQPLPAKLRDNLARPDA
SEGLQALADNAWVSEFLPLIPGVDPVPGQARLSVGQORVAVARALNPSLIL
LDEPAASLDHASEQVWMEALNALSRLTTLMTHTQLEDLADNDVIMWQDRIIEQGR
YALISVAGFPATLAAHQEST"
/complement(5083..6048)
/gene="trxB"
/locus_tag="S0888"
/complement(5083..6048)
/gene="trxB"
/locus_tag="S0888"
/function="enzyme; 2-Decoxyribonucleotide metabolism"
/notes="residues 1 to 321 of 321 are 95.95 pct identical to residues 1 to 321 of 321 from Escherichia coli K-12 : B0888"
/codon_start=1
/transl_table=11
/product="thioredoxin reductase"
/protein_id="AAP16352.1"
/db_xref="GI:30040621"
/translation="MGTTKSKLILGSGPYTAAYAAARANLOPVILTGMKGGOL
TTTTEVNMEDPDNDLGPILMARMEHATKFEFEIIFDHINKYDLQNRPRLANGNG
EYTDALITATGASARYLGPSEBAFGRGVSACTCDGFFYRQKVAIVGGATYAV
EALYLSINIASVHMLHRRDGRRAKILIKRLMDVENVGNIIILHNRKLEVTGQMGV
TGVRLRDTQNSDNIESLDVAGLFVAIGHSPTALFEQLELENGYIKVQSGIHGNATQ
TSIPGVPAAGVDMDHIVRAITSAIGTCMALDAERYLIDGLADAK"
/6592..7086
/gene="trp"
/locus_tag="S0889"
/6592..7086
/gene="trp"
/locus_tag="S0889"
/function="regulator; Central intermediary metabolism: Pool, multipurpose conversions"
/notes="residues 1 to 164 of 164 are 100.00 pct identical to residues 1 to 164 of 164 from Escherichia coli K-12 : B0889"
/codon_start=1
/transl_table=11
/product="regulator for leucine (or trp) regulon and high-affinity branched-chain amino acid transport system"
/protein_id="AAP16353.1"
/db_xref="GI:30040622"
/translation="MVDSKRPGLDRIIDRIINLEOKDGRISNVELSKRVGSPFP
CLEVRRLRQSGFIQGTALLNPNYDASLILVPEITLNGAPVPFQPOFNAVOKLE
IOERHLVSGPDIYILKTRVDMXRYKRLIGSETLRLRGVNDRTYVMEVKQSNRLV
IKTR"
/7221..11249
/gene="ftrK"
/locus_tag="S0890"
/7221..11249
/gene="ftrK"
/locus_tag="S0890"
/function="phenotype; Cell division"
/notes="residues 1 to 737 of 1342 are 68.24 pct identical to residues 1 to 737 of 1339 from Escherichia coli K-12 : B0890"

```

```

/codon_start=1
/transl_table=1
/product="cell division protein"
/protein_id="AAP16354.1"
/db_xref="GI:30040623"
/translation="MSGEYEDKEVLTLLTSSGRLLLEALLTLVLPVAMMALISF
NPSPDSMTAMHEPIHNGMFGMALDTLFFIGVAVYTPVITVGGCFAMRHSS
SDEVIDEFAVSLRITIGVALILTSGLAINDIMVFAAGVIGSLSTLLOPLLS
SGGRTILLCVMAAGLTLFTGMSVVTIAELGDMINILTPASNRRTRVDDEDE
DDEEDENHNGKHESRRARILRGALRRKLAERKINMGRODALLFSKRDDDE
EITTLAKGVAADDDVLFSGNRATOEYDEYDYLINSAPITEPVAATAATQVAV
AVEPVTQTPVAVSDVPPSOPTVAMOPVGPGRGEPITAPAGSIFQSQVAPVA
VNEPLQOPVOPPYAPAPAEOPACOPYAAPREPVPVAGNMAOEEOGTFAPSTQ
TEQTOOPAEPLVQOPDUEOCVBERVEBETKAPRLYFEEVEKERAREE
OLAAWYOPVPEVKEPRKSLKAPVAAPVAAVAASPLASGVKATLTGAAA
TVAAVPSLANSGGPRPOVKEGIGPOLRPRIRVETRELASYGILPQRAAEKA
RRAORNOVSQVYNDIDELAMQDELARFOCTOQOORGEYOHDVPAVEADAAA
QPOQVAPQOQVQOQCPVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQ
APQVQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQV
DLTPPSEVEPVTFALEOMARLVEARLADFRKADLVAVSGPVITREMLAVGV
KAARISNLSRLARSLSTVAARVVEVPGKRYGELTPNKROTIVLRREVLDNAKPRD

```

```

Query Match
Best Local Similarity 98.0%; Score 340.8; DB 1; Length 292504;
Matches 345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 105 CGGTAAATATAGCGGCCCAATTCGAGCTGAACATTACAGTACGCGGTAACTC 164
DB 212560 CAGTAAATATAGCGGCCCAATTCAGAGCTGAACATTACAGTACGCGGTAACTC 212619
QY 165 TGCACCTGCTGCAAACTGATGCCCGTAACTGACTGATTAATCCAGCATGCGCG 224
DB 212620 TGCACCTGCTGCAAACTGATGCCCGTAACTGACTGATTAATCCAGCATGCGCG 212679
QY 225 CGGTAAATGTCAGATTTGTCAGAGGCTCAGATGACGCTAAATGATCTGACCAACG 284
DB 212680 TGGTAATGTCAGATTTGTCAGAGGCTCAGATGACGCTAAATGATCTGACCAACG 212739
QY 285 TGGCTTGGTAAAGCGGCTTACTTTGATCAGTGAACCGCAAAATTCGAAATGACGCT 344
DB 212740 TGGCTTGGTAAAGCGGCTTACTTTGATCAGTGAACCGCAAAATTCGAAATGACGCT 212799
QY 345 TAAACAGTTGCGTGTGCGCAACGCGTCTCAGTTACCAAGCTGATTAATCTCTCCGT 404
DB 212800 TAAACAGTTGCGTGTGCGCAACGCGTCTCAGTTACCAAGCTGATTAATCTCTCCGT 212859
QY 405 CAACGTGACTCAGGTGGTGGTGAACAACGCGACCGCTCATGACTAA 456
DB 212860 CAACGTGACTCAGGTGGTGGTGAACAACGCGACCGCTCATGACTAA 212911

```

```

RESULT 12
AP237726/c 437 bp DNA linear BCT 25-MAY-2000
LOCUS AP237726
DEFINITION Shigella flexneri insertion sequence IS600, complete sequence.
ACCESSION AP237726
VERSION AF237726.1 GI:7582371
KEYWORDS
SOURCE
ORGANISM Shigella flexneri
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Shigella flexneri
1 (bases 1 to 437)
Sakellaris, H., Hammink, N. K., Rajakumar, K., Bulach, D., Hunt, M.,
Sasakawa, C. and Adler, B.
TITLE Curli loci of Shigella spp
JOURNAL Infect. Immun. 68 (6), 3780-3783 (2000)
MEDLINE 20278182
PUBMED 10816548
REFERENCE 2 (bases 1 to 437)
Sakellaris, H., Hammink, N. K., Rajakumar, K., Bulach, D., Hunt, M.,
Sasakawa, C. and Adler, B.
AUTHORS

```

```

TITLE Direct Submission
JOURNAL Submitted (23-FEB-2000) Microbiology, Monash University, Wellington
Rd., Clayton, Vic 3800, Australia
FEATURES
source
1..437
/organism="Shigella flexneri"
/mol_type="genomic DNA"
/strain="SBA1100"
/serotype="2a"
/db_xref="taxon:623"
320..392
/organism="Shigella flexneri"
/mol_type="genomic DNA"
/strain="SBA1100"
/serotype="2a"
/db_xref="taxon:623"
/insertion_seq="IS600"
/notice="similar to IS600 deposited in GenBank Accession
Number X05952"
join(1..319,393..437)
/notice="csg locus; similar to Escherichia coli K12 csg
locus deposited in GenBank Accession Number X90754"
misc_feature
ORIGIN

```

```

Query Match
Best Local Similarity 98.1%; Score 311.4; DB 1; Length 437;
Matches 315; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 105 CGGTAAATATAGCGGCCCAATTCGAGCTGAACATTACAGTACGCGGTAACTC 164
DB 321 CAGTAAATATAGCGGCCCAATTCAGAGCTGAACATTACAGTACGCGGTAACTC 262
QY 165 TGCACCTGCTGCAAACTGATGCCCGTAACTGACTGATTAATCCAGCATGCGCG 224
DB 261 TGCACCTGCTGCAAACTGATGCCCGTAACTGACTGATTAATCCAGCATGCGCG 202
QY 225 CGGTAAATGTCAGATTTGTCAGAGGCTCAGATGACGCTAAATGATCTGACCAACG 284
DB 201 TGGTAATGTCAGATTTGTCAGAGGCTCAGATGACGCTAAATGATCTGACCAACG 142
QY 285 TGGCTTGGTAAAGCGGCTTACTTTGATCAGTGAACCGCAAAATTCGAAATGACGCT 344
DB 141 TGGCTTGGTAAAGCGGCTTACTTTGATCAGTGAACCGCAAAATTCGAAATGACGCT 82
QY 345 TAAACAGTTGCGTGTGCGCAACGCGTCTCAGTTACCAAGCTGATTAATCTCTCCGT 404
DB 81 TAAACAGTTGCGTGTGCGCAACGCGTCTCAGTTACCAAGCTGATTAATCTCTCCGT 22
QY 405 CAACGTGACTCAGGTGGTGGTGAACAACGCGACCGCTCATGACTAA 425
DB 21 CAACGTGACTCAGGTGGTGGTGAACAACGCGACCGCTCATGACTAA 1

```

```

RESULT 13
STAJ2301 5103 bp DNA linear BCT 15-NOV-2000
LOCUS STAJ2301
DEFINITION Salmonella typhimurium csgC, csgF, csgB, csgD, csgB, csgA, and c
ACCESSION AJ002301
VERSION AJ002301.1 GI:2739232
KEYWORDS
SOURCE
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1
Romling, U., Bian, Z., Hammink, M., Sieralta, W. D. and Normark, S.
Curli fibers are highly conserved between Salmonella typhimurium
and Escherichia coli with respect to operon structure and
regulation
JOURNAL J. Bacteriol. 180 (3), 722-731 (1998)
MEDLINE 98117058

```

PUBMED 9457880
REFERENCE 2 (bases 1 to 5103)
AUTHORS Romling U.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES
SOURCE
1. 5103
Location/Qualifiers
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="SR-11"
/db_xref="taxon:602"
1. 3322
/note="ORF179a"
/codon_start=2
/transl_table=1
/product="hypothetical protein"
/protein_id="CA05311.1"
/db_xref="GI:2739233"
/db_xref="GOA:O54290"
/db_xref="SWISS-PROT:O54290"
/translation="CTLLSGVMALVIIHSGALAPHLIEIVSVLTGIIVAFIMCIQANQ
LILSPVDTFGACATFAGQDMRLVLPALGLIFGYAMKNSGIMLASRREQHSANT
AVTK"
complement(361. 1194)
/gene="cs9G"
complement(361. 1194)
/gene="cs9G"
/function="1,10protease"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="cs9G protein"
/protein_id="CA05312.1"
/db_xref="GI:2739234"
/db_xref="GOA:O54291"
/db_xref="SWISS-PROT:O54291"
/translation="MPRLILVAVLLISGCTAPPKQAQKPTLMPRAQSYKDLTHLP
PTGKIPFVSIVYIODETGQFKFYPASNSTAPQSAIHALVKDSRFTPLEQSGQ
NLNLRKIIIRAQENGTAVAMNRIPLQSLTANLVESGIIGESNVKSGGVARYG
IGADTQYQDQIAVNLRVNVSTGEIISVNTSKTIIISYEVOAGVFRIIDYQRLLEG
IGYTSNEPVMCLMSALETGVIPLINDIGRLMDLQWKADRONDIIVKRELSVPE
S"
complement(1221. 1637)
/gene="cs9F"
complement(1221. 1637)
/gene="cs9F"
/codon_start=1
/transl_table=1
/product="cs9F protein"
/protein_id="CA05313.1"
/db_xref="GI:2739235"
/db_xref="SWISS-PROT:O54292"
/translation="MRVYHAVVLLMFLSPPLWAGNMTFQFRNPNFGGNPNNGSFLNS
AQANSYKDPAYNDFGIETPSALDNFOAIQSOILGGLLTNITGKRGAVNTDFII
DIARBDQLOLVNTRKTRGPRSTIEVSGLOIQTDP"
complement(1664. 2059)
/gene="cs9E"
complement(1664. 2059)
/gene="cs9E"
/codon_start=1
/transl_table=1
/product="cs9E protein"
/protein_id="CA05314.1"
/db_xref="GI:2739236"
/db_xref="SWISS-PROT:O54293"
/translation="MKRYLTWIVAAILPATGNLHANEVEVEPGILTDHTVSGIHE
FYRPSDKWESEYGNLTINERPARSGSWTTIVNDVIFQTLFPMKDFEKTIVF
ALAQTERALNRQIDQTLSTSDLARDEF"
complement(2064. 2714)
/gene="cs9D"
complement(2064. 2714)
/gene="cs9D"
complement(2064. 2714)
/gene="cs9D"

/function="transcriptional regulator"
/codon_start=1
/transl_table=1
/product="CsgD protein"
/protein_id="CA05315.1"
/db_xref="GI:2739237"
/db_xref="GOA:O54294"
/db_xref="SWISS-PROT:O54294"
/translation="WFNEVSSHGHTLLITKPSLQATFALIQHLKOSLATITGKLNIIQ
RSLIEDISAGCIVLMDMEADKLIHYQDNLNRKNNIKTLLNTPDPYPREIENMP
HINGVFATQEDQEHVSGIQILRGECYFSQKLASVLIHSGNRYNSTESALLTHRE
KEILNKLRIGTSSNEIRSLFISENTVKTLYNLFKKIAYKGRFOAVSWANDNRR"
complement(2869. 2896)
/gene="cs9DERG"
complement(2869. 2874)
/gene="cs9DERG"
complement(2891. 2896)
/gene="cs9DERG"
3344. 3376
/gene="cs9BA"
3344. 3349
/gene="cs9BA"
3371. 3376
/gene="cs9BA"
3469. 3924
/gene="cs9B"
3469. 3924
/gene="cs9B"
3469. 3924
/gene="cs9B"
/function="nuclease"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="CsgB protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translation="MKRKLPMHLTIIGAIGIATATYTDLARSRYNRAVELSSRN
QAATIGQVGTDSNARVQESKLSVTSQSGNNRRAVDAGNTNFAIETGTANADA
SISQAVGNSAAIIQKSGKNKANTQYGTAVVQKSHMARVQR"
3469. 3531
/gene="cs9B"
3532. 3921
/gene="cs9B"
/product="CsgB protein"
/function="nuclease"
3966. 4421
/gene="cs9A"
3966. 4421
/gene="cs9A"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="cs9A protein"
/protein_id="CA05317.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/db_xref="SWISS-PROT:P55225"
/translation="MKLIKVAFAIIVSGSALAGVVPQMGGGGNHNGGNSGPDST
LSIYQVGSAAALALQSDARKSETTITQSGYGNQADVGQGANSTTEILTQNGRNNAT
IDQMANNSDITVQIGRNALVNOTASDSVWVROVGFNNATANQY"
3966. 4025
/gene="cs9A"
4026. 4418
/gene="cs9A"
/product="cs9A protein"
/function="fiber subunit"
4483. 4809
/gene="cs9C"
4483. 4809
/gene="cs9C"
/codon_start=1
sig_peptide
mat_peptide
sig_peptide
mat_peptide
gene
CDS
sig_peptide
mat_peptide
gene
CDS

```

/transl_table=11
/product="CsgC protein"
/protein_id="CA05318.1"
/db_xref="GI:2739240"
/db_xref="SWISS-PROT:P55227"
/translation="MTLLLLAALSNQTFPTTQGGDIYVVPQTLNBPCCQVQIHLI
SVADGGGGSHTQKQKLTSLPANOPIELSLRLSNVLSSEBSKVLITVSDGSLHSQC
WPSAQ"
4483..4533
/gene="csgC"
4534..4806
/gene="csgC"
/product="CsgC"
4940..5103
/gene="ORF103"
4940..5103
/gene="ORF103"
/codon_start=1
/transl_table=11
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:O54295"
/translation="MSYIKKNIPALICLCAFFIHSAVGGQTVGGVIERGAIVPEL
CDISTHAEIN"

```

ORIGIN

Query Match	56.84;	Score 259.2;	DB 1;	Length 5103;
Best Local Similarity	73.04;	Pred. No. 2.9e-64;		
Matches 333; Conservative	0;	Mismatches 122;		

QY	1	ATGAACTTTTAAAAAGTAAAGCAATTGCAGCAATCGTATNTCTCCGGTACGCTCTGGCA	60
Db	3966	ATGAAACTTTTAAAAAGTGGCAGCATTTCCAGCAATCGTATNTCTCCGGTACGCTCTGGCT	4025
QY	61	GGATGTTCTCTAGTACAGCGCGCGCGGTTAACCAAGTGTGTGGCGTAAATTAATGACGAC	120
Db	4026	GGGTCGTTCCAAATGGGCGCGCGCGGTTAATCAATACGGCGCGCATATGTTCCGGC	4085
QY	121	CCAAATTCTGAGCTGAACATTTAACAGTACGGTGGGGTAACTGTCACCTGCTCTGCA	180
Db	4086	CCGAAATTCACGTTGGAGCATTTATCGATACGGTTCGCTACGCTGCGCTGCTCTGCA	4145
QY	181	ACGATGCGCCGATCTGACTCTGACTATTAATCCCAAGTCAGCGCGCGGTAATGATGACAGAT	240
Db	4146	AGCGATGCCCTGTAATCTGTAAACGACCAATTAACCCAGACGGTTATGTGTAACGGCGCCGAT	4205
QY	241	GTTTGTCAAGGCTCAGATGACAGCTCAATTCATCTGACCCCAAGTGGCTTCGGTAACAGC	300
Db	4206	GTAGGCGAGGGCGCGGATTAACAGTACTATTTGACTGACTCAGATGGTTTCAGAAACAAT	4265
QY	301	GCTACTCTGATCACTGGAAACGGCAAAAATTTGAAATGACGGTTAAACAGTTCCGTTGGT	360
Db	4266	GCCACCATCGAACCAAGTGAACGCTTAAAAATCTCCGATATTAATCTCGCTCAATACGGCGGT	4325
QY	361	GGCAACGGTGTGCTCAGTTGACCAAGTCAATCTAACTCTCCGCTCAACGTAATCAGGTT	420
Db	4326	AATTAACGCGCGGTGTTAATCAACGCAATCTGATTTCCAGCGCTTAATGTGCTCAGGTT	4385
QY	421	GGCTTTGGTAAACAAGGAGCAGCGCTCATCAGTACTAA	456
Db	4386	GGTTTGGCAACAACGCAACGCTTAACCAATATTTAA	4421

RESULT 14	AE008749	LOCUS	DEFINITION	ACCESION
	AE008749	22411 bp	DNA linear	BCT 23-APR-2003
			<i>Salmonella typhimurium</i> LT2, section 53 of the complete genome.	

REFERENCE
AUTHORS
1 (bases 1 to 22411) *Salmonella*.
McLelland, M., Sanderson, K. E., Speich, J., Clifton, S. W.,
Latrellle, P., Courtney, J., Potwollik, S., Ali, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulroney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
Stonking, T., Nshan, M., Waterston, R. and Wilson, R. K.
TITLE
Complete genome sequence of *Salmonella enterica* serovar Typhimurium
L12
JOURNAL
NATURE 413 (6858), 852-856 (2001)
MEDLINE
21534948
PUBMED
11677609
REFERENCE
2 (bases 1 to 22411)
TITLE
The *Salmonella typhimurium* Genome Sequencing Project
JOURNAL
Direct Submission
Submitted (23-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT
Supported by NIH grant 5U 01 AI43793

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs GLIMMER; <http://www.tigr.org/scrablab/glimmer/glimmer.html> and Genemark; <http://opal.phylogeny.gatech.edu/Genemark/>. EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.org/ecocyc/>.

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ResnombB;
http://kinich.cifn.unam.mx:8850/db/ram1/orth.htm

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30) ; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

Source

Source

1.22411

cyphimurium LT2"

```
/strain="LT2; SGSC 1412; ATCC 700720"
```

```

'xref-acc': 100120"
/db xref="taxon:99287"

```

```
/note="LT2"
```

```
/gene="phoH
```

```

424 /note="syno

```

```
/gene="rhoH
```

```

/ note="putative RBS for
454 1208

```

/geɪnə= "rhon"

```
/note="81ml"
region comm
```

starvation

add, 928.1de
/codon start

/transl_tabl

```
product="phoB-dependent ATP-binding pho regulon
component"
```

/protein_id=

/translation

RSRKQREESVTSLVQMGVEAIGMARDSRDTSPVKARNEAOAHYI.NA

YDVLKRLGASFMOCYCLRPETGKY

CDS

RBS

გენე

LTR|LBAENVY|VNGDIT|QCGL|P|GVRSG|SLD|ALR|FEE|DEN|VGI|V|H|NK|D|CVR|S|AL|S|
 ORFL|TAYNTV"
 gene complement(1414. .2303)
 /gene="STM1127"
 cds complement(1414. .2295)
 /gene="STM1127"
 /note="similar to E. coli orf, hypothetical protein
 (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
 identity in aa 15 - 264"
 /codon_start=1
 /transl_table=1
 /product="putative transcriptional regulator"
 /protein_id="AAL20058.1"
 /db_xref="GI:16419643"
 /translation="MEPPQPRFKPEKGIIDLTGAMQKSLTRASQRIAYIILAPPROVTO
 SSIALISBEKTAQGEATVIRFCRTLYGKQFQDFKNDLILATITSDSDSLAEVSE
 SDDAHLIGETIKONTISNLTSETLLMDQVQGVADMLRHCHSVYIIGVSSGSIAD
 MHKMRIGLRGDADSNHPRFMVQMTLTKADVMAGSHSGTSPETHSILRLRQGA
 TTVATLTHNIGSLPCEADPCLINGRQGMLOGDSITGRKAQDLPFDDLITLLVQSSSE
 QARESKLATTNALLDPTK"
 RBS complement(2298. .2303)
 /gene="STM1127"
 /note="putative RBS for STM1127; RegulonDB:STMSIH001399"
 complement(2580. .4088)
 /gene="STM1128"
 complement(2580. .4076)
 /gene="STM1128"
 /note="similar to E. coli putative cotransporter
 (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
 identity in aa 7 - 478"
 /codon_start=1
 /transl_table=1
 /product="putative sodium/glucose cotransporter"
 /protein_id="AAL20059.1"
 /db_xref="GI:16419644"
 /translation="MTHSFGIVLVYLFGLVLMMLVGVYFSPROKTADYDFRGGR
 VSGMAAGSVPRATLSSITPMSIPAKAFTSPWTFIIQYLAIALPLVFFIYIPFRK
 LNVATYIETLEFRVRCRLFPASMSFMLEFHIGRIATFTVLTLARPLAIDVPIVL
 LILSWCIITVYMWGIGEVITWDVIOGLULSGSALIFIVICLKVGQGIDEIFVTOQA
 DKFPAVPHVMSWETSTVPLMIGFLEPNIIQGFASQDVQVQRYVTDSEIKETKLT
 NAKYLAUVIPVEFFAIGSALFVYOOHPOLPAGNFTGILLPLFVTEMPIVNGILIA
 ALFPAAGSISISSINSSICFNSDIYQRLSHKRTPENRKTIKALVILVAGLISNAS
 VMLVNDSEILWDAPNSLIGLMDGPGTGLFMLGIFFRRANAGSVADGIILSVITLVGA
 RYADLNFPEFVIGSLSVISGVIFAPLFAPAPPLTIDEXPEPKYTL"
 complement(4083. .4088)
 /gene="STM1128"
 /note="putative RBS for STM1128; RegulonDB:STMSIH001400"
 complement(4413. .5099)
 /gene="STM1129"
 complement(4413. .5093)
 /gene="STM1129"
 /note="similar to E. coli putative enzyme (AAC76255.1);
 Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
 226"
 /codon_start=1
 /transl_table=1
 /product="putative inner membrane protein"
 /protein_id="AAL20060.1"
 /db_xref="GI:16419645"
 /translation="WSLIARLQSVHENGGLIYSCQVPPGSPMDKEPIYAAHQAAAS
 AGAVARILEGLEENKTRVPHLSVPIIGIKKIDLGSPVRTYPLQDDALDAQAGADII
 AFDAFRRPQVTDLSLTRLRHGLLMAVDCSTYNEGISCHQKIEGIFGTLLEGYGP
 ITDPVDAIPADYQQLSHAGCRVIAEGRYVTPALANALIEHGMAMVAVGSAITRYEHICQ
 WPSHVRKR"
 complement(5094. .5099)
 /gene="STM1129"
 /note="putative RBS for STM1129; RegulonDB:STMSIH001401"
 5599. .6759
 /gene="STM1130"
 5599. .6759
 /gene="STM1130"
 /note="similar to E. coli orf, hypothetical protein
 (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%

Query Match	Best Local Similarity	Score	DB 1;	Length	22411;
Matches	333;	Conservative	0;	Mismatches	123;
				Indels	0;
				Gaps	0;
QY	1	ATGAACCTTTAAAGTAGAAGCAATTGACGCAATCGTATTC	CCGGTAGCGCTTGGCA	60	
Db	17769	ATGAACTTTTAAAGTAGAGCGACATTTGACGCAATCGTATTC	TGCAAGTCTTGGCT	1782	
QY	61	GGTGTGTTCTCTCAGTACGCGCGCGCGGCTAATTAATGACGCG		120	
Db	17829	GGCGTCGTTCCACAGTAGGGGGGGGCGGAGTATCATATACGCGCGCATATGTTCCGCG		1788	
QY	121	CCAAATTTCTGACTGAACAATTATACAGTACGCGTGGCGGTAACTCTGCACTTGGCTTGGCA		180	
Db	17889	CCGGAATTCACAGTTGACATTTATCAGTAGCGGTTCCGCTTAACCTGGCTTGGCTTGGCA		17944	
QY	181	ACTGATGCGCGTAACTCTGACCTTGAACATTAACCCAGATGCGCGCGGCTTAATGATGACAGAT		240	
Db	17949	AGCGATGCGCGTAACTCTGAAACCACTTAACCCAGAGCGGCTTAATGATGATGACGCGCGAT		18000	
QY	241	GTGTGTCAGGGCTCAGATGACAGCTCAATCGATCTGACCCAAAGTGGCTTCGGTAAACAGC		300	
Db	18009	GTAGGCGACGGGTGGCGATTAACAGTACATTAATGAATGACTCAGATGGTTTCAAAACAAAT		18061	
QY	301	GTACTCTTGATCAGTGAACCGCAAAATTTCTGAATGACGGTTAAACAGTTTCGGTGT		360	


```
/note="Orthologue of E. coli P75863; Faeta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:O82787"
/db_xref="SPTREMBL:Q82787"
/translation="MATHSLRFLHPVSKMRGIGLTHALADISGLAFDRIFWVTSDDT
PITARQPMWFTPSPLHDGLHLTPAGSSALVAFDTPTQDAFTVWGNHPTARVA
PTAINOMLSGFSRDLVOLCAWGPOLTRVKNHNAVPLGADGYPYLINBSLNDLOQ
RCPAGVQWEOFRPNLVVGVAAWEDSKVIRIGVIRIDVVKPSGRCIPTTVSPKGO
KHPSGEPLATQAFRTADNGDVQGVLTARNSGVIRVGEVEHLIARAKAGATT
LDSTVPERKPDGSTTTMOCGTGCGNNQVVLBELENGIRIPISCPAGICGCCRIK
LLEBSVSPKSAAGDDSTILSCSVPTALRLN"
/complement(4737, .4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
/gene="STY1082"
/translation="PS01261 Uncharacterized protein family UPF0020
signature"
/signature="
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
overlap"
misc_feature
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
5506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
/signature="
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
overlap"
CDS
Query Match 56.5%; Score 257.6; DB 1; Length 254050;
Best Local Similarity 72.8%; Pred.No.1.5e-63;
Matches 332; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
```

```
DB 88906 ATGAACTTTTAAAGTGAAGCAATTCGACCAATCGTAGTTTCGAGAGCTCTGACT 88965
QY 61 GGTGTTGTCCTCAGTACGGCGGCGCGGTAAACAGGTGCGGTAAATAGCGGC 120
DB 88966 GCGGCGTTCACATAGGGGGGGGGGGGTATATACAGCGCGGCAATAGTTCCGGG 89025
QY 121 CCAATTCTGAGCTGAACATTATACAGTACGGTGGCGGTAACTCTGCACTTCTCTGCA 180
DB 89026 CCGGATTCACAGTTGAGCATTTATACAGTACGGTTCGGTAAACGTCGCTCTCTGCA 89085
QY 181 ACTATGCCGTAACTCTGACTTAACTTACCCAGATGGCGGGGTATGTCAGAT 240
DB 89086 AGCGATGCCCTTAATCTGAAACGACCAATCCAGACGGTTATGTAACGGCGGAT 89145
QY 241 GTTGATCAGGGCTCAGATGACGATCAATCATCTGACCCAGCTGCGTCTGTAACAGC 300
DB 89146 GTAGGCCAGGGTGGGATTAACAGTACTATTGAACTGACTGAAATGGTTTCAGAAACAT 89205
QY 301 GCTACTCTTATCATGTGGAACGGCAAAATTTCTGAATGACGTTAAACAGTTGCGTGT 360
DB 89206 GCCACCAATCGACAGGTGGAACGTAAACCTCCGATATTACTGTCGTCATTAACGGCGGT 89265
QY 361 GGCAACGGTGTGCGAGTTGACCAAGATGATCTTAACTCCGTCACAGTACTGAGGT 420
DB 89266 AATAACCGCGCGCTGTTAATCAACCGCATCTTATTCACCGTAAATGTCGTCAGGT 89325
QY 421 GCGTTTGTAAACAGCGACCGCTCATCATGACTTAA 456
DB 89326 GTTTGGCAACAGCGCGGCTTAACAGTATTAA 89361
```

Search completed: March 15, 2004, 22:49:48
Job time : 1966.17 secs

QY 1 ATGAACTTTTAAAGTGAAGCAATTCGACCAATCGTAGTTTCGAGAGCTCTGCA 60

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-3

Perfect score: 456

Sequence: 1 atgaacttttaaaagtaga.....cgaccgcctacagctactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database:

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	254.4	55.8	456	1	US-08-233-788A-58
2	173.4	38.0	361	1	US-08-233-788A-56
3	35.6	7.8	1788	4	US-09-252-991A-4632
4	35.6	7.8	2250	4	US-09-252-991A-4181
5	35.6	7.8	3102	4	US-09-252-991A-4429
6	34	7.5	552	4	US-09-252-991A-2024
7	34	7.5	669	4	US-09-252-991A-1852
8	34	7.5	1185	4	US-09-252-991A-1931
9	33.4	7.3	4403765	3	US-09-103-840A-2
10	33.4	7.3	4411529	3	US-09-103-840A-1
11	33	7.2	1830121	4	US-08-557-884-1
12	33	7.2	1830121	4	US-08-643-990A-1
13	32	7.0	2993	2	US-08-225-488-6
14	31.8	6.9	4403765	3	US-09-103-840A-2
15	31.4	6.9	1317	4	US-09-252-991A-10989
16	31.2	6.8	1810	4	US-09-667-135-29
17	31	6.8	1422	4	US-09-328-352-586
18	30.6	6.7	435	3	US-08-569-147-75
19	30.6	6.7	1923	4	US-09-056-556-190
20	30.6	6.7	1923	4	US-09-072-596-185
21	30.6	6.7	1923	4	US-09-072-596-190
22	30.6	6.7	4411529	3	US-09-103-840A-1
23	30	6.6	6122	1	US-08-403-545-1
24	30	6.6	6122	3	US-08-404-381-1
25	30	6.6	1230025	4	US-09-198-452A-1
26	29.8	6.5	771	4	US-09-489-039A-915
27	29.8	6.5	1128	4	US-09-328-352-3993

28	29.8	6.5	2652	4	US-09-489-039A-5584	Sequence 584, Ap
29	29.8	6.5	87563	4	US-09-453-702B-57	Sequence 57, Appl
30	29.2	6.4	14654	4	US-08-961-577-106	Sequence 106, App
31	28.8	6.3	429	4	US-09-252-991A-8302	Sequence 8302, Ap
32	28.8	6.3	912	2	US-08-993-228-7	Sequence 7, Appl1
33	28.8	6.3	1023	4	US-09-328-352-1184	Sequence 3184, Ap
34	28.8	6.3	1395	4	US-09-252-991A-8312	Sequence 8312, Ap
35	28.8	6.3	9507	4	US-09-489-039A-6896	Sequence 6896, Ap
36	28.8	6.3	12720	1	US-08-403-866-11	Sequence 11, Appl
37	28.6	6.3	217	4	US-09-313-294A-5247	Sequence 5247, Ap
38	28.6	6.3	693	4	US-09-328-352-3008	Sequence 3008, Ap
39	28.6	6.3	1507	4	US-09-453-323-1	Sequence 9, Appl1
40	28.6	6.3	1873	3	US-09-461-474-9	Sequence 2442, Ap
41	28.4	6.2	1380	4	US-09-489-039A-2442	Sequence 32, Appl
42	28.4	6.2	62909	4	US-09-596-002-32	Sequence 4713, Ap
43	28.2	6.2	298	4	US-09-313-294A-4713	Sequence 317, App
44	28.2	6.2	346	3	US-09-060-756-317	Sequence 317, App
45	28.2	6.2	346	4	US-09-670-314-317	Sequence 317, App

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Sharon S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
TITLE OF INVENTION: OF SALMONELLA
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 55.8%; Score 254.4; DB 1; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.3e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0;

QY 1 ATGAACCTTTAAAGTAGAAGCAATTCGAGCAATCTGTCGGTAGGCGCTTGCA 60
 DB 1 ATGAACCTTTAAAGTAGAAGCAATTCGAGCAATCTGTCGGTAGGCGCTTGCA 60
 QY 61 GGTGTTGTTCTCAGTAGAGCGCGCGCGGTAAACGAGTGGCGGTAAATTAACGGC 120
 DB 61 GGTGTTGTTCTCAGTAGAGCGCGCGCGGTAAACGAGTGGCGGTAAATTAACGGC 120
 QY 121 CCAATTTCTAGCTGAACATTTACCAATTCGAGTGGCGGTAACTCTGACTGCTGCA 180
 DB 121 CCGAGCTCAAGCTTAGCATTTATCATAGATGCTCCCTTACGCTGCTGCTGCA 180
 QY 181 ACTGATGCGCGGTAACTGACTGACTTAATCCAGAGTGGCGCGGTAAATGCTCAAT 240
 DB 181 ACGGATGCGCGGTAACTGACTGACTTAATCCAGAGTGGCGGTAAATGCTCAAT 240
 QY 241 GTTGTGTCAGGCTCAGATGACAGCTCAATGATCTGACCCAGAGTGGCGGTAAACGC 300
 DB 241 GTAGGCGCGGTGCGGTAAATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 301 GCTACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 DB 301 GCTACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 QY 361 GCGAAGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 361 AATGATGCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 QY 421 GCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 456
 DB 421 GCGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 456

RESULT 2

US-08-233-788A-56
 / Sequence 56, Application US/08233788A
 / Patent No. 5635617
 / GENERAL INFORMATION:
 / APPLICANT: Doran, James L.
 / APPLICANT: Kay, William W.
 / APPLICANT: Collinson, Karen S.
 / APPLICANT: Clouthier, Sharon C.
 / TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 / NUMBER OF SEQUENCES: 61
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Seed and Berry
 / STREET: 6300 Columbia Center, 701 Fifth Avenue
 / CITY: Seattle
 / STATE: Washington
 / COUNTRY: U.S.A.
 / ZIP: 98104-7092
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patent In Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/233,788A
 / FILING DATE: 26-APR-1994
 / CLASSIFICATION: 435
 / ATTORNEY/AGENT INFORMATION:
 / NAME: King, Joshua
 / REGISTRATION NUMBER: 35,570
 / REFERENCE/DOCKET NUMBER: 920043.403C2
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (206) 622-4900
 / TELEFAX: (206) 682-6031
 / TELEX: 372836 SEEDANBERY
 / INFORMATION FOR SEQ ID NO: 56:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 361 base pairs

TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..357
 US-08-233-788A-56

Query Match 38.0%; Score 173.4; DB 1; Length 361;
 Best Local Similarity 69.9%; Pred. No. 7.3e-45;
 Matches 234; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 64 GTTGTCTCTGATGAGCGCGCGGTAAACGAGTGGCGGTAAATTAACGGCCCA 123
 DB 1 GTGTTACCAAGTGGCGCGCGGTAAACGAGTGGCGGTAAATTAACGGCCCA 123
 QY 124 AATTGTAGCTGAACATTTACCAATTCGAGTGGCGGTAACTCTGACTGCTGCA 183
 DB 61 GACTCAACGTTGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 183
 QY 184 GATGCGCGGTAACTGACTGACTTAATCCAGAGTGGCGCGGTAAATGCTGCAAT 243
 DB 121 GATGCGCGGTAACTGACTGACTTAATCCAGAGTGGCGCGGTAAATGCTGCAAT 243
 QY 244 GGTGAGGCTCAGATGACAGCTCAATGATGATGATGATGATGATGATGATGAT 303
 DB 181 GGTGAGGCTCAGATGACAGCTCAATGATGATGATGATGATGATGATGATGAT 303
 QY 304 ACTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
 DB 241 ACCATGACAGTGGCGCGGTAAACGAGTGGCGGTAAATTAACGGCCCA 363
 QY 364 AACGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
 DB 301 AACGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420

RESULT 3

US-09-252-991A-4632/C
 / Sequence 4632, Application US/09252991A
 / Patent No. 6551795
 / GENERAL INFORMATION:
 / APPLICANT: Marc J. Rubinfeld et al.
 / TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 / FILE REFERENCE: 107196.136
 / CURRENT APPLICATION NUMBER: US/09/252,991A
 / CURRENT FILING DATE: 1999-02-18
 / PRIOR APPLICATION NUMBER: US 60/074,788
 / PRIOR FILING DATE: 1998-02-18
 / PRIOR APPLICATION NUMBER: US 60/094,190
 / PRIOR FILING DATE: 1998-07-27
 / NUMBER OF SEQ ID NOS: 33142
 / SEQ ID NO 4632
 / LENGTH: 1788
 / TYPE: DNA
 / ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-4632

Query Match 7.8%; Score 35.6; DB 4; Length 1788;
 Best Local Similarity 64.6%; Pred. No. 0.23;
 Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 30 AGCAATGATATTCCTCGTAGAGCTCTGCGAGTGTCTCTCAAGTACGCGCGCGCG 89
 DB 1391 AGCGGTGATATTCCTCGTAGAGCTCTGCGAGTGTCTCTCAAGTACGCGCGCGCG 89
 QY 90 TAACACGCGTGTGCGGTAT 111
 DB 1331 TGATCAGGTGCGGTAT 1310

RESULT 4
US-09-252-991A-4181
; Sequence 4181, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4181
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4181

Query Match 7.8%; Score 35.6; DB 4; Length 2250;
Best Local Similarity 64.6%; Pred. No. 0.26;
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Qy 30 AGCAATGATTTCTCCGAGCTCTGCGAGTGTCTCTCACTACGCGCGCGCGG 89
Db 339 AGCGGCTATTTCGCGGTCGCGAGCTCGCGGTGGCATCGGAGGCGCGCGCGG 398
Qy 90 TAACCAAGGTGTGGCGGTAAT 111
Db 399 TGATCAGGTGCGCGGTGAGCAT 420

RESULT 5
US-09-252-991A-4429/c
; Sequence 4429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4429
; LENGTH: 3102
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4429

Query Match 7.8%; Score 35.6; DB 4; Length 3102;
Best Local Similarity 64.6%; Pred. No. 0.3;
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Qy 30 AGCAATGATTTCTCCGAGCTCTGCGAGTGTCTCTCACTACGCGCGCGCGG 89
Db 1411 AGCGGCTATTTCGCGGTCGCGAGCTCGCGGTGGCATCGGAGGCGCGCGG 1352
Qy 90 TAACCAAGGTGTGGCGGTAAT 111
Db 1351 TGATCAGGTGCGCGGTGAGCAT 1330

RESULT 6
US-09-252-991A-2024/c
; Sequence 2024, Application US/09252991A

Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2024
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2024

Query Match 7.5%; Score 34; DB 4; Length 552;
Best Local Similarity 56.1%; Pred. No. 0.44;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 147 GTACGTTGCGGTAAGTCTGCTGCTGCACTGATGCGCGTAAGTCTGACTTAC 206
Db 119 GTGCGGCGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 60
Qy 207 TATTACCAAGCATGCGCGGTAATGTGTCAGATGTGTCAGGAGCTCAATGA 260
Db 59 TGTGCTCTTGTGCGCGGTGTCATGATGTCGCGGCTGCGGTCAGAGGTCTGCTGA 6

RESULT 7
US-09-252-991A-1852
; Sequence 1852, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1852
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1852

Query Match 7.5%; Score 34; DB 4; Length 669;
Best Local Similarity 56.1%; Pred. No. 0.48;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 147 GTACGTTGCGGTAAGTCTGCTGCTGCACTGATGCGCGTAAGTCTGACTTAC 206
Db 462 GTGCGGCGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 521
Qy 207 TATTACCAAGCATGCGCGGTAATGTGTCAGATGTGTCAGGAGCTCAATGA 260
Db 522 TGTGCTCTTGTGCGCGGTGTCATGATGTCGCGGCTGCGGTCAGAGGTCTGCTGA 575

RESULT 8
US-09-252-991A-1931
; Sequence 1931, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; PRIOR FILING DATE: 1999-02-18
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO: 1931
;; LENGTH: 1185
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1931

Query Match
Best Local Similarity 7.5%; Score 34; DB 4; Length 1185;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 147 GTACCGTGGCGGTACTCTGCACTTGTCTGCAAACTGATGCCCGTAACTGACTTGAC 206
Db 782 GTGCGGGGGCGCGCTGTCTGTTCGCGCGCTCCGATGCCGCGCTCTGCTGCTCC 841
Qy 207 TATATCCAGCATGCGCGCGTAAATGTCGATGTTGTGTCAGGCGCTCAGATGA 260
Db 842 TGTTCCTTGTGCGCGCTGTTCATGTGTGCGCGCTGGGTCAGGGTCTGCTGA 895

RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO: 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match
Best Local Similarity 7.3%; Score 33.4; DB 3; Length 4403765;
Matches 43; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 206 CTATTACCCAGCATGCGCGCGTAAATGTCGATGTTGTGTCAGGCGCTCAGATGAAC 264
Db 371246 CTGTGCGGAGCAGCGCGCGGATGTCGATGTTGTGCGCGCTCAGATGAAGGC 371304

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.

;; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;; FILE REFERENCE: 24366-20007.00
;; CURRENT APPLICATION NUMBER: US/09/103,840A
;; PRIOR FILING DATE: 1998-06-24
;; NUMBER OF SEQ ID NOS: 2
;; SOFTWARE: Patent in Ver. 2.1
;; SEQ ID NO: 1
;; LENGTH: 4411529
;; TYPE: DNA
;; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match
Best Local Similarity 7.3%; Score 33.4; DB 3; Length 4411529;
Matches 43; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 206 CTATTACCCAGCATGCGCGCGTAAATGTCGATGTTGTGTCAGGCGCTCAGATGAAC 264
Db 371189 CTGTGCGGAGCAGCGCGGATGTCGATGTTGTGCGCGCTCAGATGAAGGC 371247

RESULT 11
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match
Best Local Similarity 7.2%; Score 33; DB 4; Length 1830121;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 122 CAATTCGAGCTAATCTTACCATGAGTGGCGGTAACTCTGCACTTGCTGCAAA 181

Db 740925 CGATTTTCGCAAGATCGTTTAAATGATATGCGCTTACAAATTCCTTGTTATGACACA 740984

Qy 182 CTGATGCCCGTAACTCTGACTTGTATTTACCCAGCATGGCGCGGTAAATGCT 234

Db 740985 TTGATGCTGTGTTTTTAAACCAAAATTATTTGAAATGTCGGCGTTAAGGCT 741037

RESULT 12

US-09-643-990A-1

Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams

Owen White
Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-AUG-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match 7.2%; Score 33; DB 4; Length 1830121;

Best Local Similarity 55.8%; Pred. No. 30;

Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 122 CAAATTCGAGCTGAACATTTTACCAAGTCGGTGGGGAATCTGCACTGCTGCAAA 181

Db 740925 CGATTTTCGCAAGATCGTTTAAATGATATGCGCTTACAAATTCCTTGTTATGACACA 740984

Qy 182 CTGATGCCCGTAACTCTGACTTGTATTTACCCAGCATGGCGCGGTAAATGCT 234

Db 740985 TTGATGCTGTGTTTTTAAACCAAAATTATTTGAAATGTCGGCGTTAAGGCT 741037

RESULT 13

US-08-225-488-6/c
Sequence 6, Application US/08225488

Patent No. 5846802

GENERAL INFORMATION:

APPLICANT: Buxton, Frank

APPLICANT: Hinnen, Albert

APPLICANT: Visser, Jacob

TITLE OF INVENTION: No. 5846802el Fungal Protease

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/225,488

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/047,214

FILING DATE: 13-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Spruill, W. Murray

REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: 4-19055/A/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8615

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2993 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Aspergillus niger

STRAIN: M400

IMMEDIATE SOURCE:

CLONE: pTZPBDP

FEATURE:

NAME/KEY: promoter

LOCATION: 1..829

FEATURE:

NAME/KEY: CDS

LOCATION: join(830..1153, 1205..1649, 1697..1785, 1841

LOCATION: 2233)

OTHER INFORMATION: /note="subtilisin-type protease

OTHER INFORMATION: PEPP of Aspergillus niger; product of gene pepd"

FEATURE:

NAME/KEY: intron

LOCATION: 1154..1204

FEATURE:

NAME/KEY: intron

LOCATION: 1650..1696

FEATURE:

NAME/KEY: intron

LOCATION: 1786..1840

US-08-225-488-6

Query Match 7.0%; Score 32; DB 2; Length 2993;

Best Local Similarity 46.4%; Pred. No. 3.9;

Matches 104; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 65 TTGTTCTCAGTACGGGGCGGCGGTAACCAAGGTGTGGCTTAATTAATAGCGCCCA 124
|||
Db 2181 TTGTAGCCGAGAAATTGGGGCTAACCGGCAATGTGTACAGCATTTCCGCTAGCGCAAC 2122
QY 125 ATTCTGAGCTGAACATTTACCAAGTACGGTGGCGGTAAGTCTGTGCACTTGTGCAAACTG 184
|||
Db 2121 CTCTTAGCTCGGTGCTGTGAGCGGCTGGGGTATGCAAGTCCCGCAAGCCCATCAAAATAG 2062
QY 185 ATGCCCGTAACTGTGACTTATTAACCAAGTACGGGCGGTAATGGTGCAGATGTTG 244
|||
Db 2061 AGGATTAACCTGTGCAATGAGGTGAGCCATGACGTCCCGAGATCGTGTGGTGCC 2002
QY 245 GTCAGGCTCAGATGACGTCAATGATCTGACCAACGATGCG 288
|||
Db 2001 GAGTTGAGCGGCTCATGACAGAAATCTTGTCTTCCCGGCGC 1958
|||

RESULT 14

US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

Query Match 7.0%; Score 31.8; DB 3; Length 4403765;
Best Local Similarity 46.0%; Pred. No. 72;

Matches 108; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 82 GCGCGCGGTAAACCGGTGTGGGTATATATAGCGCCCAATTCTGAGCTGAACATT 141
|||
Db 4029096 GCGGTGGGGGTGACGCTGTGAGGCGGTAAACGCGGATCGCGGTGACGCGCGCGC 4029037
QY 142 TACCAAGTACGGTGGCTTAATCTGCACTTGTGCAACTGCAATGCGCGTAATCTGAC 201
|||
Db 4029036 GCGGAGACGCGCGCGCGCGCGCGCGCGCGCGCTTTCGCGCTTCCGTAAGCTCCGG 4028977
QY 202 TTGACTATATACCGACATGCGCGCGGTATATGTGTGACAGATTTGTCAGAGGCTCAGATGAC 261
|||
Db 4028976 CTGGCGCGTGGCGCGGTAGCGGGGTATATGCGCGCGCGGTGTGAGCCCGGTGTGGCC 4028917
QY 262 AGCTCAATGATCTGACCAACGTGCTTGGTAAACGCGCTACTTGTATGATGCT 316
|||
Db 4028916 GCGTCCCGCGCGGTGGGCTTGGCGGCGGTGGCGCGCAACCTCGCGCAGT 4028862
|||

RESULT 15

US-09-252-991A-10989/c
; Sequence 10989, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136

;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 10989
;; LENGTH: 1317
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10989

Query Match 6.9%; Score 31.4; DB 4; Length 1317;
Best Local Similarity 52.7%; Pred. No. 4.2;

Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 64 GTTGTCTCAGTACCGCGCGGCGGTAACCAAGGTGTGGCGGTAAATTAATAGCGCCCA 123
|||
Db 513 GTTGTCTCAGTACCGCGCGGCGGCGGCGGCGGTGTGTGGAGGCGGACGTCGATCGCTC 454
QY 124 AATTCTGAGCTGAACATTTACCAAGTACGGTGGCGGTAATCTGTGCACTTGTGCAAACT 183
|||
Db 453 GCGACTGCGGAGCGCTTGTCCAGTTCTTGGCGGTGGGAGGTGTGTACTGACGCG 394
QY 184 GATGCCCGT 192
|||
Db 393 GATGCCCGT 385
|||

Search completed: March 16, 2004, 04:36:54
Job time: 72.6647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-3

Sequence: 1 atgaaactttaagtaga.....cgaccgctcatcagtactaa 456

Scoring table: IDENTITY_NUC

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Maximum DB seq length: 20000000000

Post-processing: Minimum Match of

Listing first 45 summaries

Database : Published Applications NA: *

- 1: /cgn2_6/prodata/2/pubnna/US07_PUBCOMB.seq;
- 2: /cgn2_6/prodata/2/pubnna/PC1_NEW_PUB.seq;
- 3: /cgn2_6/prodata/2/pubnna/US06_NEW_PUB.seq;
- 4: /cgn2_6/prodata/2/pubnna/US06_PUBCOMB.seq;
- 5: /cgn2_6/prodata/2/pubnna/US07_NEW_PUB.seq;
- 6: /cgn2_6/prodata/2/pubnna/PC1US_PUBCOMB.seq;
- 7: /cgn2_6/prodata/2/pubnna/US08_NEW_PUB.seq;
- 8: /cgn2_6/prodata/2/pubnna/US08_PUBCOMB.seq;
- 9: /cgn2_6/prodata/2/pubnna/US09A_PUBCOMB.seq;
- 10: /cgn2_6/prodata/2/pubnna/US09B_PUBCOMB.seq;
- 11: /cgn2_6/prodata/2/pubnna/US09C_PUBCOMB.seq;
- 12: /cgn2_6/prodata/2/pubnna/US09_NEW_PUB.seq;
- 13: /cgn2_6/prodata/2/pubnna/US10A_PUBCOMB.seq;
- 14: /cgn2_6/prodata/2/pubnna/US10C_PUBCOMB.seq;
- 15: /cgn2_6/prodata/2/pubnna/US10D_PUBCOMB.seq;
- 16: /cgn2_6/prodata/2/pubnna/US10_NEW_PUB.seq;
- 17: /cgn2_6/prodata/2/pubnna/US60_NEW_PUB.seq;
- 18: /cgn2_6/prodata/2/pubnna/US60_PUBCOMB.seq;

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Query				Description	
Result No.	Score	Match length	ID		
c 1	42.6	9.3	13529	12	US-10-249-445-1585
	39	8.6	1350	14	US-10-017-161-1781
	33	8.6	1350	15	US-10-292-789-1437
c 3	38.2	8.4	635	14	US-10-029-386-24675
c 5	37.4	8.2	1972	14	US-10-017-161-1981
c 6	37.4	8.2	1972	15	US-10-293-798-1629
c 7	37	8.1	2560	9	US-09-738-626-1857
c 8	37	8.1	3309400	9	US-09-738-626-1
c 9	35.8	7.9	1493	14	US-10-029-386-25133
c 10	35.6	7.8	992	12	US-10-425-114-31747
c 11	35.6	7.8	2122	15	US-10-106-8204-1242
c 12	35.6	7.8	3078	12	US-10-389-687-726
c 13	35.6	7.8	4389	15	US-10-368-493-66788
c 14	35	7.7	785	14	US-10-028-386-22627
c 15	34.6	7.6	532	14	US-10-029-386-7202

C	16	34.6	7.6	1633	14	US-10-017-161-1819	Sequence 1475, Ap
C	17	34.6	7.6	1633	15	US-10-392-1998-1475	Sequence 1475, Ap
C	18	34.6	7.6	155074	13	US-10-026-188-6	Sequence 6, Appl
C	19	34.2	7.5	528	14	US-10-029-386-5350	Sequence 5350, Ap
C	20	34.2	7.5	666	14	US-10-029-386-25948	Sequence 25948, Ap
C	21	34.2	7.5	1632	14	US-10-355-430-33	Sequence 33, Appl
C	22	34.2	7.5	3592	14	US-10-017-161-1913	Sequence 1913, Ap
C	23	34.2	7.5	3532	15	US-10-292-798-1569	Sequence 1569, Ap
C	24	34.2	7.5	12981	15	US-10-369-493-33865	Sequence 33865, Ap
C	25	34.2	7.5	33918	14	US-10-017-161-2049	Sequence 2049, Ap
C	26	34.2	7.5	33918	15	US-10-292-798-1695	Sequence 1695, Ap
C	27	34	7.5	570	9	US-09-864-761-21778	Sequence 21778, A
C	28	34	7.5	1976	9	US-09-864-761-5052	Sequence 5052, Ap
C	29	33.8	7.4	765	15	US-10-369-493-44503	Sequence 44503, Ap
C	30	33.6	7.4	652	15	US-10-260-238-4681	Sequence 4681, A
C	31	33.6	7.4	1197	14	US-10-156-761-4579	Sequence 4579, Ap
C	32	33.6	7.4	9025608	14	US-10-156-761-1	Sequence 1, Appl
C	33	33.4	7.3	5163	12	US-10-282-1228-26364	Sequence 26364, A
C	34	33.4	7.3	6615	12	US-10-282-1228-28180	Sequence 28180, A
C	35	33.2	7.3	832	12	US-10-282-1228-19206	Sequence 19206, A
C	36	33.2	7.3	1377	14	US-10-338-075-1261	Sequence 1261, Ap
C	37	33.2	7.3	11475	14	US-10-238-075-1250	Sequence 1250, Ap
C	38	33	7.2	379	14	US-10-029-386-19106	Sequence 19106, A
C	39	33	7.2	569	14	US-10-029-386-3710	Sequence 3710, Ap
C	40	33	7.2	652	14	US-10-029-386-37283	Sequence 37283, A
C	41	33	7.2	867	12	US-10-282-1228-19705	Sequence 19705, A
C	42	33	7.2	1959	9	US-09-864-761-4012	Sequence 4012, Ap
C	43	33	7.2	1973	9	US-09-864-761-3471	Sequence 3471, Ap
C	44	33	7.2	3897	12	US-10-282-1228-22081	Sequence 22081, A
C	45	33	7.2	9590	12	US-10-282-1228-33351	Sequence 33351, A

ALIGNMENTS

RESULT 1

```

# Sequence 1585, Application US/10240425
# Publication No. US20040033502A1
# GENERAL INFORMATION:
# APPLICANT: Williams, Amanda
# APPLICANT: Boland, Joseph F.
# APPLICANT: Lord, Reginald V.
# APPLICANT: Alvarez, Chris
# APPLICANT: Wetzel, Jon C.
# APPLICANT: Scheif, Uwe
# APPLICANT: Vockley, Joseph G.
# TITLE OF INVENTION: Gene Expression Profiles in Esophageal Tissue
# FILE REFERENCE: 44921-5026
# CURRENT APPLICATION NUMBER: US/10/240,425
# CURRENT FILING DATE: 2002-09-30
# PRIOR APPLICATION NUMBER: PCT/US01/09847
# PRIOR FILING DATE: 2001-03-28
# PRIOR APPLICATION NUMBER: US 60/193,446
# PRIOR FILING DATE: 2000-03-31
# NUMBER OF SEQ ID NOS: 1588
# SOFTWARE: Patentin Ver. 2.1
# SEQ ID NO 1585
# LENGTH: 135259
# TYPE: DNA
# ORGANISM: Homo sapiens
# FEATURE:
# OTHER INFORMATION: Genbank Accession No. US20040033502A1 Z83838
# US-10-240-425-1585
```

Query Match	Score	DB	Length
9.34	42.6	12	135259

Matches 160; Conservative 0; Mismatches 174; Indels 3; Gaps 1

Oy 47 GTATGCGCTCTCCAGAGTGTTCCTCCAGTACCGGCGCGCGGTAAACACAGGTGGTGGCG 106
 Db 70242 GTAATGGTGGTGGTGTATGGTATGGGTATGATGGTGGTGGCTGTATCACAGTATGATGTG 70183

```

Qy 107 GTAATATAGGCGCCAAATTCTGAGTGAACATTACAGTACGTCGCGTAACCTG 166
Db 70182 GTGGTATGATGATCATGATGATGATGATGATGATGATGATGATGATGATG 166
Qy 167 CACTTGCTCTGCAAACTGATGATGATGATGATGATGATGATGATGATGATG 70123
Db 70122 ATGGTCGACACCAAGATGATGATGATGATGATGATGATGATGATGATG 226
Qy 227 GTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70066
Db 70065 ATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 286
Qy 287 GCTTCGCTGACACCGCTACTCTTATGATGATGATGATGATGATGATGATG 70006
Db 70005 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 346
Qy 347 AACAGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69946
Db 69945 GTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69909

```

RESULT 2

```

US-10-017-161-1781
; Sequence 1781, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABRABANT, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; PRIOR FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1781
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-017-161-1781

```

Query Match
Best Local Similarity 8.6%; Score 39; DB 14; Length 1390;
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```

Qy 62 GTGTTGTTCTCAGTACGCGCGCGGTAACACGCTGATGCGGTATTAATACGGCC 121
Db 853 GTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
Qy 122 CAAATTCGAGCTGAACATTACACGATGCGTACCTGCACTTCTGCAAA 181
Db 913 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 972
Qy 182 CTGATGCGCGTAACCTGATGATGATGATGATGATGATGATGATGATGATG 241
Db 973 GTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1032
Qy 242 TTGGTCAGGCTCAGATGACAGCTCAATGATCTGACCAAGCGTTCGTAACGCG 301
Db 1033 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
Qy 302 CTACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
Db 1093 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1152

```

```

Qy 362 GCAACGCTGCTGACGTTGA 380
Db 1153 GCAATGCTGATGATGATG 1171

```

RESULT 3

```

US-10-292-798-1437
; Sequence 1437, Application US/10292798
; Publication No. US2003023583A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABRABANT, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1437
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-292-798-1437

```

Query Match
Best Local Similarity 8.6%; Score 39; DB 15; Length 1390;
Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```

Qy 62 GTGTTGTTCTCAGTACGCGCGCGGTAACACGCTGATGCGGTATTAATACGGCC 121
Db 853 GTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
Qy 122 CAAATTCGAGCTGAACATTACACGATGCGTACCTGCACTTCTGCAAA 181
Db 913 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 972
Qy 182 CTGATGCGCGTAACCTGATGATGATGATGATGATGATGATGATGATGATG 241
Db 973 GTGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1032
Qy 242 TTGGTCAGGCTCAGATGACAGCTCAATGATCTGACCAAGCGTTCGTAACGCG 301
Db 1033 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
Qy 302 CTACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
Db 1093 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1152
Qy 362 GCAACGCTGCTGACGTTGA 380
Db 1153 GCAATGCTGATGATGATG 1171

```

RESULT 4

```

US-10-029-386-24675
; Sequence 24675, Application US/10029386
; Publication No. US2003019470A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.

```

```
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
FILE REFERENCE: AEMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24675
LENGTH: 625
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP001754.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
OTHER INFORMATION: EST HUMAN HIT: B1015471.1, EVALUATE 1.00e-03
US-10-029-386-24675

Query Match      8.4%; Score 38.2; DB 14; Length 625;
Best Local Similarity 46.2%; Pred. No. 0.026;
Matches 162; Conservative 0; Mismatches 188; Indels 1; Gaps 1;

QY 29 CAGCAGTGTATCTCCGATAGCGCTCTGCGACAGTGTCTTCTCAGTACGCGCGCGCG 88
DB 87 CATCCCTTACTTATGAGTAGTGGGCGCGGTGGTGGTATGATGATGATGATGATGATG 146
QY 89 GTAACACAGGCGTGGCGGCTATATAGCGCCCAATTCAGAGTGAACATTACCGCT 148
DB 147 ATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 205
QY 149 ACGGTGCGGCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 208
DB 206 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 265
QY 209 TTACCCAGCAGTGGCGCGGCTATGATGATGATGATGATGATGATGATGATGATG 268
DB 266 ATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 325
QY 269 TCGATCTGACCCCAAGTGGCTTCCGTACACCGCTGCTGCTGCTGCTGCTGCTGCTG 328
DB 326 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 385
QY 329 ATTCTGAATGACGCTTAAACAGTTCGGTGGGCAACGCTGCTGCTGCTGCTGCTG 379
DB 386 ATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 436

RESULT 5
US-10-017-161-1981/c
Sequence 1981, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1981
LENGTH: 1972
TYPE: DNA
ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1972)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(1772)
FEATURE:
NAME/KEY: modified base
LOCATION: (975)..(994)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (1319)..(1328)
OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1981

Query Match      8.2%; Score 37.4; DB 14; Length 1972;
Best Local Similarity 49.2%; Pred. No. 0.085;
Matches 98; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 62 GTGTTGTTCTCAGTACGCGCGCGGCTAACACAGGTGGTGGCGGTAAATATAGCGGCC 121
DB 460 GTGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 401
QY 122 CAATTCGAGCTGAACATTACAGTACGCTGGCGGTAATCTGCACTTCTGCAAA 181
DB 400 ATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 341
QY 182 CTGATGCGCCGTAATCTGACTTGAATTAACCCAGCAGTGGCGCGGTAATGTCAGATG 241
DB 340 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 281
QY 242 TTGCTCAGCGCTCAGATGA 260
DB 280 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 262

RESULT 6
US-10-292-798-1629/c
Sequence 1629, Application US/10292798
Publication No. US2003023583A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 084335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1629
LENGTH: 1972
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: source
FEATURE:
LOCATION: (1)..(1972)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(1772)
FEATURE:
NAME/KEY: modified base
LOCATION: (975)..(994)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
```

LOCATION: (1319)..(1328)
 OTHER INFORMATION: a, c, g, unknown or other
 US-10-292-798-1629

Query Match
 Best Local Similarity 8.2%; Score 37.4; DB 15; Length 1972;
 Matches 98; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 62 GTGTGTCTCCAGTACGCGCGCGGTACCAACCGGTGTGTGGCGGTATATATACGGCC 121
 DB 460 GTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 121
 QY 122 CAATTTTGAAGCTGAACATTATACAGTGTGGCGGTACTCTGCTCTGCAAA 181
 DB 400 ATG 341
 QY 182 CTGATGCCGTACTCTGACTTACTATTAACCAAGCGCGGTATATGATGATGATG 241
 DB 340 GTGATG 281
 QY 242 TTGTGTAGGCTTCAGATGA 260
 DB 280 GTG 262

RESULT 7

US-09-738-626-1857
 Sequence 1857, Application US/09738626
 Publication No. US20020197605A1
 GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
 APPLICANT: MIZOGUCHI, HIROSHI
 APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIAI, KEIKO
 APPLICANT: YOKOI, HARUHIKO
 APPLICANT: TATEISHI, NAOKO
 APPLICANT: SENOH, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738, 626
 PRIOR FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 1857
 LENGTH: 2580
 TYPE: DNA

ORGANISM: Corynebacterium glutamicum
 US-09-738-626-1857

Query Match
 Best Local Similarity 8.1%; Score 37; DB 9; Length 2580;
 Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 77 ACGCGCGCGCGGTAAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 136
 DB 1046 AGGTGTAGCGT 1105
 QY 137 ACATTTTACAGT 1105
 DB 1106 ACACGACCAAGT 1165
 QY 194 ACTGTGACTTGTATTAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 253
 DB 1166 CCATTAATTAAGGT 1225

QY 254 CAGATCAGACGTCAATTCATCTGACCCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 313
 DB 1226 CAGATGTGCGAGTCCATTTACGATTCATGAAACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1285

QY 314 AGTGAAGCGGCAAAATTTCTGAATGACGTTAAACAGTTTGTGTGTGTGTGTGTGTGTGTGT 366
 DB 1286 TGT 1338

RESULT 8

US-09-738-626-1/c
 Sequence 1, Application US/09738626
 Publication No. US20020197605A1
 GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
 APPLICANT: MIZOGUCHI, HIROSHI
 APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIAI, KEIKO
 APPLICANT: YOKOI, HARUHIKO
 APPLICANT: TATEISHI, NAOKO
 APPLICANT: SENOH, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738, 626
 PRIOR FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 1
 LENGTH: 3309400
 TYPE: DNA

ORGANISM: Corynebacterium glutamicum
 US-09-738-626-1

Query Match
 Best Local Similarity 8.1%; Score 37; DB 9; Length 3309400;
 Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 77 ACGCGCGCGCGGTAAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 136
 DB 1784428 AGGTGTAGCGT 1784369
 QY 137 ACATTTTACAGT 193
 DB 1784368 ACACGACCAAGT 1784309
 QY 194 ACTGTGACTTGTATTAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1784309
 DB 1784308 CCATTAATTAAGGT 1784249
 QY 254 CAGATGACGCTCAATTCATCTGACCCAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 313
 DB 1784248 CAGATGTGCGAGTCCATTTACGATTCATGAAACCGTGTGTGTGTGTGTGTGTGTGT 1784189
 QY 314 AGTGAAGCGGCAAAATTTCTGAATGACGTTAAACAGTTTGTGTGTGTGTGTGTGTGTGTGT 366
 DB 1784188 TGT 1784136

RESULT 9

US-10-029-386-25133/c
 Sequence 25133, Application US/10029386
 Publication No. US20030194704A1
 GENERAL INFORMATION:

Query Match	7.9%	Score 35.8	DB 14	Length 1493
Best Local Similarity	44.2%	Pred. No. 0.26		
Matches 148; Conservative	0	Mismatches 187	Indels 0	Gaps 0

RESULT 10
 US-10-425-114-31747/c
 : Sequence 31747, Application US/10425114
 : Publication No. US2004003488A1
 : GENERAL INFORMATION:
 : APPLICANT: Liu, Jingdong
 : APPLICANT: Zhou, Yihua
 : APPLICANT: Kovalic, David K.
 : APPLICANT: Screen, Steven E
 : APPLICANT: Tabaska, Jack E
 : APPLICANT: Cao, Yongwei
 : TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 : TITLE OF INVENTION: Plants and Use Thereof for Plant Improvement
 : FILE REFERENCE: 38-21(53313)B
 : CURRENT APPLICATION NUMBER: US/10/425, 114
 : CURRENT FILING DATE: 2003-04-28
 : NUMBER OF SEQ ID NOS: 73128
 : SEQ ID NO 31747

```

US-10-425-114-31747
Query Match          7.8%; Score 35.6; DB 12; Length 992;
Best Local Similarity 54.6%; Pred. No. 0.25;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY      45 CGGTAGCGCTTGGCAGGTGTTGTTCTTCCTCACTACGCGCGCGCGGTAAACAACGCTGCTGG 104
        |||
DB       471 CCGGTACAGTTCGACCGCCCTTGATGTCTCGGCGCGCGCGCGCCCTTGGCGGTGGCGG 412

QY      105 CGGTAAATTAATAGCGGCCCAAATTCTGAGCTGAACATTTAACAGTAGACGGTGGCGGTAACTC 164
        |||
DB       411 CGGCTTGCTGTGGCGGCTCAAGATCGGACTGTGGCGGCGAAGCGGTGGCCCTGGCGGTGGCAG 352

QY      165 TGCACCTTGCT 174
        |||
DB       351 TGGGCTCGGT 342

RESULT 11
US-10-108-260A-1242
; Sequence 1242, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1242
; LENGTH: 2122
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1242

Query Match          7.8%; Score 35.6; DB 15; Length 2122;
Best Local Similarity 46.7%; Pred. No. 0.37;
Matches 113; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY      90 TAACCAAGGTGTGGCGGTAAATAATAGCGGCCCAAATTTGAGCTGAACATTTACCAGTA 149
        |||
DB       302 TATCCAAGATTCGAAGGCTAGTAATTGGTGAAGAAAATTTGCAAGAACAATTCCTCCAGTG 361

QY      150 CGGTGGGAGGTAACTCTGCACCTTGCTGCAAGCACTGATGCGCCGTAACTCTGACTGACTAT 209
        |||
DB       362 GTGAGACTGTTTTGTCTTCCTTCAGCCCGGAAAGCAAGACTGCTCCCTCTTTATTTG 421

QY      210 TACCAGCATGCGCGCGGTATGTGTGACGATGTTGTGACGGGCTCAGATGACAGCTCAAT 269
        |||
DB       422 CAAGCTCTAAACCATCAAGCAAGCAAGCAAGCACTGCAAGGCGCGGACCTTCCTGAGA 481

QY      270 CGATCTGACCAACGCTGGCTTCGGTAACAGGCTACTCTTGATCAAGTGGAACGGCAAAAA 329
        |||
DB       482 CGCACCCCAAGCAAGTGAGGAAGAACACTTGGCGCACTGTGAGGCGTTGGAAGGAGGAACA 541

QY      330 TT 331
        |
DB       542 CT 543

RESULT 12
US-10-389-647-226/c
; Sequence 226, Application US/10389647
; Publication No. US2004003549A1
; GENERAL INFORMATION:
; APPLICANT: GREENBERG, E. Peter
; APPLICANT: SCHUSTER, Martin
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
; FILE REFERENCE: UIZ-038CP
; CURRENT APPLICATION NUMBER: US//10/389,647

```

```

RESULT 15
US-10-029-386-7202
; Sequence 7202, Application US/10029386
; Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
TITLE OF INVENTION: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR G
FILE REFERENCE: AEWICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 7202
LENGTH: 532
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC003693.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-3

Perfect score: 456

Sequence: 1 atgaacttttaaaagtaga.....cgaccgcctacagctactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_estcl:*
10: gb_estcl2:*
11: gb_hrc:*
12: gb_estcl3:*
13: gb_estcl4:*
14: gb_estcl5:*
15: em_estlum:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rtd:*
26: em_ges_phg:*
27: em_ges_vrt1:*
28: gb_gescl:*
29: gb_gescl2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	453.4	99.4	680 12	BU618688 BU618688
2	60.8	13.3	558 14	CD374421 TNNBmC2H
3	43.4	9.5	607 12	BU348812 BU348812
4	38	8.3	282 29	AL758610 AL758610 Arabidops

C 5	37.2	8.2	540	12	BM382763
C 6	37.2	8.2	771	28	B2997124
C 7	37.2	8.2	797	28	B2997127
C 8	37.2	8.2	818	29	CG200536
C 9	37	8.1	789	28	B2115869
C 10	36	7.9	1171	13	BX342087
C 11	35.8	7.9	675	14	CG642466
C 12	35.8	7.9	774	24	CNS02466
C 13	35.8	7.9	935	10	CNS02466
C 14	35.6	7.8	435	10	BE604165
C 15	35.6	7.8	498	10	BE604467
C 16	35.6	7.8	523	12	BM380389
C 17	35.6	7.8	537	28	B64870
C 18	35.6	7.8	642	28	B2431645
C 19	35.6	7.8	652	12	BM379271
C 20	35.6	7.8	684	29	CG16249
C 21	35.6	7.8	769	29	CG101505
C 22	35.6	7.8	793	29	CG364848
C 23	35.6	7.8	839	28	BZ969045
C 24	35.6	7.8	854	29	CG142634
C 25	35.6	7.8	867	29	CG142632
C 26	35.6	7.8	880	28	BZ571017
C 27	35.6	7.8	921	29	CG364838
C 28	35.6	7.8	926	29	CG627745
C 29	35.6	7.8	926	29	CG211571
C 30	35.6	7.8	950	28	CC327866
C 31	35.6	7.8	970	29	CG651748
C 32	35.6	7.8	1245	28	BZ571017
C 33	35.4	7.8	244	29	CE219585
C 34	35.2	7.7	379	12	BZ66281
C 35	35.2	7.7	634	12	BZ66096
C 36	35.2	7.7	708	12	BZ722995
C 37	35.2	7.7	1251	14	CD509711
C 38	35	7.7	629	12	BU404131
C 39	34.8	7.6	685	12	BU249080
C 40	34.6	7.6	582	12	BM276856
C 41	34.6	7.6	586	10	AM759708
C 42	34.6	7.6	701	14	CD436924
C 43	34.6	7.6	818	14	CD038848
C 44	34.6	7.6	978	29	CNS00K6G
C 45	34.2	7.5	306	14	R55792

ALIGNMENTS

RESULT 1
LOCUS BU618688/c 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BU618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis CDNA clone XL186b22 5', mRNA sequence.
ACCESSION BU618688
VERSION BU618688.1 GI:37256713
KEYWORDS
SOURCE
ORANISM Xenopus laevis (African clawed frog)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.

REFERENCE
AUTHORS Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and Kohara, Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
CONTACT: Tadashi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genie.nig.ac.jp
URL: The information of this clone is available through the following

URL: The information of this clone is available through the following

http://xenopus.nibb.ac.jp.
 Location/Qualifiers
 1.680
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="XL186b22"
 /issue_type="whole embryo"
 /dev_stage="stage 10.5"
 /clone_lib="NIBB Mochi normalized Xenopus early gastrula library"

Query Match
 Best Local Similarity 99.4%; Score 453.4; DB 12; Length 680;
 Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Db 1 ATGAACCTTTAAAGTAAAGCAATTCAGACATCGATTCCTCGGTAGCCCTCTGGCA 60
 593 ATGAACCTTTAAAGTAAAGCAATTCAGACATCGATTCCTCGGTAGCCCTCTGGCA 60
 61 GGTGTTCTCTCAAGTACGGCGCGCGGTAAACCAAGCGGTGGGTAAATAGCGGC 474
 533 GGTGTTCTCTCAAGTACGGCGCGCGGTAAACCAAGCGGTGGGTAAATAGCGGC 474
 121 CCAATTCCTAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGACCTTCTGCA 180
 473 CCAATTCCTAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGACCTTCTGCA 180
 181 ACTGATCCCGTAACTCTGACTTGAATTAACCCAGCAATGGCGCGGTAAATAGTGCAGAT 240
 413 ACTGATCCCGTAACTCTGACTTGAATTAACCCAGCAATGGCGCGGTAAATAGTGCAGAT 240
 241 GTTGTGACGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354
 353 GTTGTGACGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354
 301 GCTACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 294
 293 GCTACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 361 GGTAAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234
 233 GGTAAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 421 GGTAAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 174
 173 GGTAAACCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 138

RESULT 2
 CD374421/c 558 bp mRNA linear EST 19-AUG-2003
 LOCUS TNNMfC2H33 Wuchereria bancrofti microfilaria cDNA (SAM9551L-WbMf)
 DEFINITION Wuchereria bancrofti cDNA clone TNNMfC2H3 5', mRNA sequence.
 ACCESSION CD374421
 VERSION CD374421.1 GI:31229939
 KEYWORDS EST.
 SOURCE Wuchereria bancrofti
 ORGANISM Wuchereria bancrofti
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Wuchereria.
 1 (bases 1 to 558)
 Nd1,J.S., Ribeiro,J.M. and Nutman,T.B.
 Toward the characterization of the Wuchereria bancrofti
 microfilarial transcriptome with comparisons to those of Brugia
 malayi and Onchocerca volvulus
 Unpublished (2003)
 Contact: Thomas B. Nutman
 Laboratory of Parasitic Diseases
 NIH
 Building 4 Room 126, Bethesda, MD 20892-0425, USA
 Email: tnutman@niaid.nih.gov

Seq primer: pJunescript T3.
 Location/Qualifiers
 1.558
 /organism="Wuchereria bancrofti"
 /mol_type="mRNA"
 /db_xref="taxon:6293"
 /clone="TNNMfC2H3"
 /dev_stage="microfilaria"
 /lab_host="XLI-Blue MRP"
 /clone_lib="Wuchereria bancrofti microfilaria cDNA (SAM9551L-WbMf)"
 /note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2: Xho I; Lymphatic filarial nematode parasite of humans, isolated from the blood of an infected individual from Guyana, South America and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 2.9 x 10⁵ independent recombinants and the average insert size is ~1kb. The library was constructed by Sandra J. Laney. The library is available from Dr.S.A.Williams, email: genome@smith.edu."

Query Match
 Best Local Similarity 13.3%; Score 60.8; DB 14; Length 558;
 Matches 93; Conservative 0; Mismatches 25; Indels 2; Gaps 2;
 Db 63 TGTGTTCTCTCAAGTACGGCGCGGTAAACCAAGCGGTGGGTAAATAGCGGC 121
 174 TGTGTTCTCTCAAGTACGGCGCGGTAAACCAAGCGGTGGGTAAATAGCGGC 115
 122 CAAATTCCTAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGACCTTCTGCA 180
 114 CAAATTCCTAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGACCTTCTGCA 55

RESULT 3
 B3348812/c 607 bp mRNA linear EST 06-MAR-2002
 LOCUS B3348812 Dictyostelium discoideum cDNA library, Af Dictyostelium
 DEFINITION B3348812 Dictyostelium discoideum cDNA clone dda34011 3', mRNA sequence.
 ACCESSION B3348812
 VERSION B3348812.1 GI:19219319
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 1 (bases 1 to 607)
 Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
 Full length cDNA of Dictyostelium discoideum at the aggregation
 stage
 Unpublished (2002)
 Contact: Tadashi Shin-i
 Center for Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.

FEATURES
 source
 1.607
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dda34011"
 /sex="mat A"
 /dev_stage="aggregation stage"
 /clone_lib="Dictyostelium discoideum cDNA library, Af"
 Query Match
 9.5%; Score 43.4; DB 12; Length 607;

```

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-160B03-013216"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/notes="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were

```

```

Email: schmable@iastate.edu
Individual basecall and confidence value were assigned using the
phred software(http://www.phrap.org/). Overall sequence quality
assessment and vector trimming was conducted using the Lucy
software (version 1.16s, http://www.tigr.org/sofft/lab/). Lucy
parameters were set to ensure an overall trimmed quality of 97.5%
or better without any vector fragments in the chosen high quality
region of each sequence. Low-quality bases between the poly-T and
the high-quality region were replaced with N's to serve as spacers
using a Perl program (est_process.pl), written by Dr. Hui-Hsien
Chou.

PCR Primers
FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: primer Sp6 (ATT TAA GTG ACA CTA TAG)
Seq primer: universal (ATT AAA CGA CGG CCA GT)
POLYA=Yes.

Location/Qualifiers
1.540
/organism="Zea mays"
/mol_type="mRNA"
/cultivar="B73"
/db_xref="taxon:4577"
/clone="MEST554-D07"
/tissue_type="mixed"
/lab_host="DH10B"

```


/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBTA381104"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"

Query Match 8.2%; Score 37.2; DB 28; Length 797;
Best Local Similarity 50.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 193 AACTTGACTGACTATTAACCCAGATGCGCGGTAATGTCGAGATGTTGTCAGGCG 252
DB 505 AACCTTTCGTGATTAACCCAGATGCGCGGTAATGTCGAGATGTTGTCAGGCGTAA 564
QY 253 TCAGATGACAGCTCATGATCGATCGACCCAGCGTTCGTTGTAACAGCGCTACTCTTGT 312
DB 565 TAATCCAAAGCTTCCACTAGTATTAACCAAGCTTTCATGCGTAATTAACCCAGCTTCAT 624
QY 313 CAGTGAACCGCAAAATTTCTGAAATGACGTTAAACAGTTGCGTGGCAACGCGTG 370
DB 625 CGGTATATATCGAACGTTTCGTGAGTAAATTAACCCAGCGCTTCGCGGTAATTAACGAG 682

RESULT 8
LOCUS CG200536 818 bp DNA linear GSS 21-AUG-2003
DEFINITION PUIUS94TD ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTA059019,
genomic survey sequence.

ACCESSION CG200536
VERSION CG200536.1 GI:34091597
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 818)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uteerback, T.,
Reinick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Benner, J.

TITLE Maize Genomes Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: 7F
Class: sheared ends.

FEATURES
SOURCE 1..818
Location/Qualifiers

/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBTA059019"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"

ORIGIN

Query Match 8.2%; Score 37.2; DB 29; Length 818;
Best Local Similarity 50.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 193 AACTTGACTGACTATTAACCCAGATGCGCGGTAATGTCGAGATGTTGTCAGGCG 252
DB 277 AACCTTTCGTGATTAACCCAGATGCGCGGTAATGTCGAGATGTTGTCAGGCGTAA 336

QY 253 TCAGATGACAGCTCATGATCGATCGACCCAGCGTTCGTTGTAACAGCGCTACTCTTGT 312
DB 337 TAATCCAAAGCTTCCACTAGTATTAACCAAGCTTTCATGCGTAATTAACCCAGCTTCAT 396
QY 313 CAGTGAACCGCAAAATTTCTGAAATGACGTTAAACAGTTGCGTGGCAACGCGTG 370
DB 397 CGGTATATATCGAACGTTTCGTGAGTAAATTAACCCAGCGCTTCGCGGTAATTAACGAG 454

RESULT 9
LOCUS B2115869/c 789 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-255D7.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-255D7, genomic survey sequence.

ACCESSION B2115869
VERSION B2115869.1 GI:23756816
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 789)
Zhao, S., Shetty, J., Shatsman, S., Tseng, G., Geer, K.,
Shvachbeyn, A., Gebregorgis, B., Overton, L., Russell, D., Chen, D.,
Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other GSS: CH230-255D7.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat330.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 255 row: D column: 7
Seq primer: 7F
Class: BAC ends.

FEATURES
SOURCE 1..789
Location/Qualifiers

/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SenHsd/MCM"
/db_xref="taxon:10116"
/clone_lib="CH230-255D7"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTAKBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SenHsd/MCM) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 8.1%; Score 37; DB 28; Length 789;
Best Local Similarity 45.8%; Pred. No. 31;
Matches 127; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 46 GTAGCGCTGAGGAGGTGTTCTCTGATGACGCGGCGGTAACACGCGTGTGTC 105
DB 680 GGTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
QY 106 GGTATATATAGCGGCCCAATTTCTGAGCTGAACATTAACAGTACGCGGCGTAACCT 165
DB 620 GTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 561

[illegible]

```

Db                               1127 SS 11.28

RESULT 11
CB642466/c                      675 bp    mRNA    linear    EST 08-APR-2003
LOCUS
DEFINITION
OSUNJB02K15.r OSUNJB Oryza sativa (japonica cultivar-group) cDNA
CB642466
VERSION
KEYWORDS
SOURCE
ORGANISM
EST:
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
1 (bases 1 to 675)
Jantsarilvarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished (2003)
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR primers
FORWARD: gta aaa cga cga cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 02 row: K column: 15
Seq primer: gga aac agc tat gac cat g.
Location/Qualifiers
1..675
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSUNJB02K15"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_id="OSUNJB"
/notes="Vector: pBluescript II KS +; Site_1: EcoRI, site_2:
XhoI; 24 hrs after inoculation with Rice Blast (Che
86061)"

ORIGIN

Query Match      7.9%; Score 35.8; DB 14; Length 675;
Best Local Similarity 59.2%; Pred. No. 63;
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

Oy 6 ACTTTAAAGTGAAGACAATGTCAGCAATCGTATTCCTCGGTAGAGCGCTCTGCAGAGTGT 65
Db 626 ACTTCCACAGGTGAAGCTGCCGTGCTGCCGCCGTCTCGAGTGAACCTCTCGGGGTGT 567
Oy 66 TGTTCCTCACTAGCAGCGCGCGCGGTAAACAACAGGTGTGTGGCGGT 108
Db 566 GCTTCGCGAGCGGTGTACGTGGCAGACCAACGCGCGGTGGCGGT 524

RESULT 12
CNS02APC/c                    774 bp    DNA    linear    GSS 01-SEP-2000
LOCUS
DEFINITION
Tetradodon nigroviridis genome survey sequence pUC-ori end of clone
22115 of library G from Tetradodon nigroviridis, genomic survey
sequence.
Accession
Version
Keywords
GSS; genome survey sequence

```


SOURCE
ORGANISM Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorphi; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE
1 Roest Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fizames, C., Winkler, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL MEDLINE
20296633
10935645

TITLE
Roest Crollius, H., Jallion, O., Dasilva, C., Ozouf-Costez, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

JOURNAL MEDLINE
20359837
10899143

REFERENCE
3 (bases 1 to 774)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES
source
1..774
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="252115"
/clone_lib="G"
/note="Genoscope sequence ID : CONG252AE08SP1-end : PUC-Or1"

ORIGIN
Query Match 7.9%; Score 35.8; DB 29; Length 774;
Best Local Similarity 48.7%; Pred. No. 67;
Matches 97; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

46 GGTAGCGCTGCGCAGTGTGTTCTCAGTACGCGCGCGGCGTAAACGCGTGTGGC 105
Db 445 GGTGTGTATGTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 386
Qy 106 GGTAAATATAGCGGCCCAATCTGAGCTGAACATTTACAGTACGTCGCGTAACTCT 165
Db 385 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 326
Qy 166 GCACCTGCTCTGCAACATGATGCCGTAACCTGACTGACTTATTAACCGACATGGCGC 225
Db 325 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 266
Qy 226 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 244
Db 265 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 247

RESULT 13
BE572585 935 bp mRNA linear EST 15-AUG-2000
LOCUS BE572585
DEFINITION 6013254.F1 NCI_CGAP_Mam6 Mus musculus cDNA clone IMAGE:3603347 5',
mRNA sequence.
ACCESSION BE572585
VERSION BE572585.1 GI:9816305

KEYWORDS
EST.
Mus musculus (house mouse)

SOURCE
ORGANISM Mus musculus
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 935)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cga@b-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Plate: LHM8792 row: a column: 12
High quality sequence stop: 679.

FEATURES
source
1..935
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3603347"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Mam6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN
Query Match 7.9%; Score 35.8; DB 10; Length 935;
Best Local Similarity 53.1%; Pred. No. 73;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

104 GCGGAATATATAGGCGCCCAATCTGAGCTGAACATTTACAGTACGTCGCGTAACT 163
Db 485 GCTGAAGAAGATCTGGCCCAAGACATCATCTTCAGATCATCTGAGACTTTGGAAAGC 544
Qy 164 CTGCACTTCTCTGCAACTGATGCCGTAACCTGACTGACTTATTAACCGACATGGCG 223
Db 545 CTTTGCTTCTCTGCAAGGATGACCGTTGTCTCAGACTTGAAGTGAATCAAGATTGGG 604
Qy 224 GCGGAATGTCGACAGTTTGGT 246
Db 605 GTGATACAGATCCAGCGCTTGT 627

RESULT 14
BE604165 435 bp mRNA linear EST 21-AUG-2000
LOCUS BE604165
DEFINITION WHE1413-1416_113 1132S wheat drought stressed leaf cDNA library
Triticum aestivum cDNA clone WHE1413-1416_113_113, mRNA sequence.
ACCESSION BE604165
VERSION BE604165.1 GI:9861435
KEYWORDS EST.
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticaceae; Triticum.
1 (bases 1 to 435)
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen H.T., Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
The structure and function of the expressed portion of the wheat

JOURNAL
COMMENT

genomes - Drought stressed leaf cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: StrataGene SK primer.

FEATURES

source

Location/Qualifiers

1..435
/organism="Triticum aestivum"
/mol_type="mRNA"
/culturivar="TAM W101"
/db_xref="taxon:4565"
/clone="WHE1413-1416_113_113"
/issue_type="leaf"
/dev_stage="Full tillering stage"
/lab_host="E. coli SOLR"
/clone_lib="Wheat drought stressed leaf cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were given a gradual
stress down to 65% and 78% RWC at Texas Tech University
(D. Zhang in HT Nguyen lab). Total RNA and poly(A) RNA
were prepared, a cDNA library was made, and the cDNA
clones were in vivo excised to give pluescript phagemids
in the T7 Close lab (Choi, Close), at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

ORIGIN

Query Match 7.8%; Score 35.6; DB 10; Length 435;
Best Local Similarity 50.0%; Pred. No. 57;

Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 134 TGAACATTACCACTAGCGGCGGTAACTGCTGCTGCAAACTGATGCCGTA 193
DB 10 TCAATTGAGCGCTATGGAATGTGAAGATTCAGTTGAATGCAATCAGCTGCCAAGA 69
QY 194 ACTCTGACTGACTATTAACCGAGCGGCGGTAAATGCGAGATGTTGTGAGGGCT 253
DB 70 AGCTGAGTTTGAATCTTAATCTTCAATAAAGCCGAGATCATGAGATGATCTTTGATG 129
QY 254 CAGATGACAGCTCAATGATCTGACCAACGCTGCTTGGTAACAGCGCTACTCTTGA 311
DB 130 GGATCAAGGTTTTTTGTAACCCGAGCATTTAGTATGTCCTGTGACCGGCACCTGCTGA 187

RESULT 15
BE604467 498 bp mRNA linear EST 21-AUG-2000
LOCUS WHE1413-1416_B07_B07ZS Wheat drought stressed leaf cDNA library
DEFINITION Triticum aestivum cDNA clone WHE1413-1416_B07_B07, mRNA sequence.
ACCESSION BE604467
VERSION BE604467.1 GI:9861825
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticaceae; Triticum.
1 (bases 1 to 498)

Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Han, P.S., Heia, C.C.,
Kang, Y., Lazo, G.R., Miller, R., Nguyen H.T., Rauech, C.J.,
Seaton, C.L., Tong, J.C. and Zhang, D.

TITLE The structure and function of the expressed portion of the wheat
genomes - Drought stressed leaf cDNA library

JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson

US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov

Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: StrataGene SK primer.

FEATURES

source

Location/Qualifiers

1..498
/organism="Triticum aestivum"
/mol_type="mRNA"
/culturivar="TAM W101"
/db_xref="taxon:4565"
/clone="WHE1413-1416_B07_B07"
/issue_type="leaf"
/dev_stage="Full tillering stage"
/lab_host="E. coli SOLR"
/clone_lib="Wheat drought stressed leaf cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were given a gradual
stress down to 65% and 78% RWC at Texas Tech University
(D. Zhang in HT Nguyen lab). Total RNA and poly(A) RNA
were prepared, a cDNA library was made, and the cDNA
clones were in vivo excised to give pluescript phagemids
in the T7 Close lab (Choi, Close), at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

ORIGIN

Query Match 7.8%; Score 35.6; DB 10; Length 498;
Best Local Similarity 50.0%; Pred. No. 61;

Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 134 TGAACATTACCACTAGCGGCGGTAACTGCTGCTGCAAACTGATGCCGTA 193
DB 10 TCAATTGAGCGCTATGGAATGTGAAGATTCAGTTGAATGCAATCAGCTGCCAAGA 69
QY 194 ACTCTGACTGACTATTAACCGAGCGGCGGTAAATGCGAGATGTTGTGAGGGCT 253
DB 70 AGCTGAGTTTGAATCTTAATCTTCAATAAAGCCGAGATCATGAGATGATCTTTGATG 129
QY 254 CAGATGACAGCTCAATGATCTGACCAACGCTGCTTGGTAACAGCGCTACTCTTGA 311
DB 130 GGATCAAGGTTTTTTGTAACCCGAGCATTTAGTATGTCCTGTGACCGGCACCTGCTGA 187

Search completed: March 16, 2004, 04:28:45
Job time : 2237.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456

Sequence: 1 atgaataacaattgttatt.....ctcgctgacacacgttaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmb1: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_scs: *
12: gb_sy: *
13: gb_un: *
14: gb_vt: *
15: gb_da: *
16: gb_fun: *
17: gb_hum: *
18: gb_in: *
19: gb_mu: *
20: gb_om: *
21: gb_ov: *
22: gb_ov: *
23: gb_ph: *
24: gb_ph: *
25: gb_pl: *
26: gb_ro: *
27: gb_scs: *
28: gb_un: *
29: gb_vt: *
30: gb_hcg: *
31: gb_hcg: *
32: gb_hcg: *
33: gb_hcg: *
34: gb_hcg: *
35: gb_hcg: *
36: gb_hcg: *
37: gb_hcg: *
38: gb_hcg: *
39: gb_hcg: *
40: gb_hcg: *
41: gb_hcg: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	4680	1	ECCSGABDG
2	456	100.0	15046	1	AB000205
3	456	100.0	15047	1	D90741
4	454.4	99.6	1711	1	AF275733
5	454.4	99.6	10190	1	AB005315
6	454.4	99.6	327773	1	AB002554
7	451.2	98.9	10370	1	AB015131
8	451.2	98.9	306358	1	AB016759
9	449.6	98.6	292504	1	AB016981
10	356.8	78.2	2889	1	CSP515700
11	332.8	73.0	2067	1	SRU43280
12	332.8	73.0	5103	1	STRJ2301
13	332.8	73.0	22411	1	AB008749
14	331.2	72.6	254050	1	AL627269
15	331.2	72.6	301983	1	AB016840
16	323	70.8	2920	1	CFR515701
17	275	60.3	1212	1	BC0131756
18	266.4	58.4	1048	1	STRGFA
19	258.2	56.6	2883	1	ESAS15702
20	51.2	11.2	301214	1	AB016786
21	49.4	10.8	246435	2	AC096956
22	47.6	10.4	253001	3	AB014834
23	43.4	9.5	234090	2	AC111458
24	43.2	9.5	182148	2	AC119000
25	43	9.4	3732	3	PFAS427
26	42.6	9.3	210554	2	AC133982
27	41.4	9.1	10709	1	AB015532
28	40.4	8.9	172742	2	AC136466
29	40.2	8.8	242510	2	AC128507
30	39.8	8.7	160460	2	AC024374
31	39.8	8.7	161536	2	AC024318
32	39.8	8.7	179693	9	AL591856
33	39.8	8.7	192016	2	AC026540
34	39.6	8.7	7218	6	166494
35	39.6	8.7	257757	3	AB014837
36	39.2	8.6	74524	9	AL512508
37	39.2	8.6	190950	2	AL359639
38	38.6	8.5	50820	1	AB014163
39	38.6	8.5	310850	1	AP005142
40	38.2	8.4	4860	6	AR320671
41	38.2	8.4	186401	2	AC097009
42	38.2	8.4	242730	2	AC099408
43	37.8	8.3	2000	6	AK655393
44	37.8	8.3	31521	3	CEY47H10A
45	37.8	8.3	141489	10	AL591376

ALIGNMENTS

RESULT 1
LOCUS ECCSGABDG
DEFINITION E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION X90754
VERSION X90754.1 GI:1147558
KEYWORDS csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
orfC gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS Hammar, M., Arnyvist, A., Bian, Z., Olesen, A. and Normark, S.
TITLE Expression of two csg operons is required for production of

Pred. No. is the number of results predicted by chance to have a

flpDnecrfin- and congo red-binding curli polymers in *Escherichia coli* K-12
 JOURNAL Mol. Microbiol. 18 (4), 661-670 (1995)
 MEDLINE 96414468
 PUBMED 8617489
 REFERENCE 2 (bases 1 to 4680)
 AUTHORS Hammar, M.
 TITLE Direct Submission
 JOURNAL Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet, Microbiology and Tumoriology Center, Box 280, S-171 77 Stockholm, SWEDEN

FEATURES
 source Location/Qualifiers
 1. 4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:562"
 /map="23.15 minutes"
 1. 3140
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="MC4100"
 /db_xref="taxon:562"
 3141.. 4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="W3110"
 /db_xref="taxon:562"
 <1..>4680
 /insertion_seq="IS2 (partial)"
 complement (1133.. 966)
 /gene="csaG"
 complement (133.. 966)
 /gene="csaG"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli production"
 /protein_id="CAA62277.1"
 /db_xref="GI:1147559"
 /db_xref="GOA:P52103"
 /db_xref="SWISS-PROT:P52103"
 /translation="MORFLVAVLLSGCUTAPPKKARPTLMPRAQSYKDTLHPA
 PFGKIFSVSYNIQDETQFQKYPASNFSTAVPOSATAMLYTALKDSKMFPIERQGLQ
 NTLNERKIRIAQENGVAIINNRIPIQSTITANIMVGSITIGESNVKSGVGARVFG
 IGADYQYLDQIAVNLRVNVSTGEITISVTSKTIIVSVQGVFRFIDYQRLLEGE
 VGYTSNEPVMCLMSAIEITGVIFILINDIGRLMDLONKARONDILVKKRHSVPE
 S"

gene
 CDS
 complement (993.. 1409)
 /gene="csaB"
 complement (993.. 1409)
 /gene="csaB"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli production"
 /protein_id="CAA62278.1"
 /db_xref="GI:1147560"
 /db_xref="SWISS-PROT:P52104"
 /translation="MRVKAIVLLMLISPLSMAGMTQFRPNRGANPNNGAFLLNS
 AQAQNSYKQPSYNDDEGIEPSPALDNFQAIOGILSLNINTGKFRGVNTDYIV
 DIANDGQLOLVNTDKTGSTIQVSGIQQNSTDF"
 complement (1434.. 1823)
 /gene="csaB"
 complement (1434.. 1823)
 /gene="csaB"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli production"
 /protein_id="CAA62279.1"
 /db_xref="GI:1147561"
 /db_xref="SWISS-PROT:P52105"

gene
 CDS
 complement (1828.. 2478)
 /gene="csaD"
 complement (1828.. 2478)
 /gene="csaD"
 /note="necessary for transcription of the csaBA operon"
 /codon_start=1
 /transl_table=11
 /product="putative regulatory protein"
 /protein_id="CAA62280.1"
 /db_xref="GI:1147562"
 /db_xref="GOA:P52106"
 /db_xref="SWISS-PROT:P52106"
 /translation="MFEVHSHIGHLTLITSSIQATALLQHLKQSLAINGKLNIQ
 RSLDDISSGSIILDMVADKGLHYWDTLSRKNNTKILNTPEDEYPRDIENNP
 HINGVPSWEDQRRVNGVQVLRGEVFTQSLASYLITSGNRYNSTESALLTHRE
 KETLAKLRIGASNNETARSLFISENTVTHLYNLPFKIAVKNRTQAVSMANDINLR"
 complement (2626)
 /note="transcriptional start site for csaDcsg operon"
 3140
 /note="transcriptional start site for csaBA operon"
 3233.. 3688
 /gene="csaB"
 3233.. 3688
 /gene="csaB"
 /codon_start=1
 /transl_table=11
 /product="nucleation component of curlin monomers"
 /protein_id="CAA62281.1"
 /db_xref="GI:1147563"
 /db_xref="GOA:P39828"
 /db_xref="SWISS-PROT:P39828"
 /translation="MKKKLFLPMVLTILGAPGIAAAGYDLANSEYFAVNEISKSFN
 OAATIGAGTNSAQLRQGGSKLAVAGQSSNRAKIDOTDYNALAYIDQGSANDA
 SISGAGVNTAMIQKSGNKNITGYQKTAIVQQRQMAIKRYTQ"
 3729.. 4184
 /gene="csaG"
 3729.. 4184
 /gene="csaG"
 /note="major subunit"
 /codon_start=1
 /transl_table=11
 /product="curlin/csaG protein"
 /protein_id="CAA62282.1"
 /db_xref="GI:1147564"
 /db_xref="GOA:P28307"
 /db_xref="SWISS-PROT:P28307"
 /translation="MKLKVAKIAIVFSGSALAGVPOYGGGNGHGGGNGSGPSE
 LIDQWNGKSEMTVKQFGGNGAAYDQTSNNSVAVTVQGFNNATIAHQ"
 3729.. 3788
 /gene="csaA"
 4243.. 4575
 /gene="orfC"
 4243.. 4575
 /gene="orfC"
 /codon_start=1
 /transl_table=11
 /product="putative curli production protein"
 /protein_id="CAA62283.1"
 /db_xref="GI:1147565"
 /db_xref="SWISS-PROT:P52107"
 /translation="MNTLILALSSQITFTNTTQGDVYTIIPVTLTQSCLCRVQIL
 SLRSGSGSQOTQKQKTLSPANOPTALTTLTSLINTSPDPRVIVTVVSPGSLHLSQO
 MPSSRSK"

ORIGIN

Query Match 100.0%; Score 456; DB 1; Length 4680;
 Best Local Similarity 100.0%; Pred. No. 2,7e-116;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 ATGAAAAACAATTGTTATTTATGATGTTAAACAATCTGGTGGCGCTGGATGGAGCC 60
Db 3233 ATGAAAAACAATTGTTATTTATGATGTTAAACAATCTGGTGGCGCTGGATGGAGCC 3292
Oy 61 GCAGCAGGTTATGATTTAGCTAATTCAGAAATTAATCTGGCGGTAAATGAATGAGTAAG 120
Db 3293 GCAGCAGGTTATGATTTAGCTAATTCAGAAATTAATCTGGCGGTAAATGAATGAGTAAG 3352
Oy 121 TCTTCATTTAATCAGGCGCCCAATATGCTCAAGCTGGAGCTCAATATATGCTCACTTA 180
Db 3353 TCTTCATTTAATCAGGCGCCCAATATGCTCAAGCTGGAGCTCAATATATGCTCACTTA 3412
Oy 181 CGCAGCAGGAGGCTCAAACTTTGGCGGTTGTTGGCAGAAAGTAGTAGCAACCGGGA 240
Db 3413 CGCAGCAGGAGGCTCAAACTTTGGCGGTTGTTGGCAGAAAGTAGTAGCAACCGGGA 3472
Oy 241 AAGATTGACACAGACAGAGATTAATTAACCTTGCAATATTTGATCAGCGCGGACGCAAC 300
Db 3473 AAGATTGACACAGACAGAGATTAATTAACCTTGCAATATTTGATCAGCGCGGACGCAAC 3532
Oy 301 GATGCGCAATTTTCCGCAAGTCTTATGTAATCTGCGATGATATATCCAGAAAGCTTCT 360
Db 3533 GATGCGCAATTTTCCGCAAGTCTTATGTAATCTGCGATGATATATCCAGAAAGCTTCT 3592
Oy 361 GGTATTAAGCAATATATTACACAGATGATGTAATCAAAAAACGCAATTGATGAGAGA 420
Db 3593 GGTATTAAGCAATATATTACACAGATGATGTAATCAAAAAACGCAATTGATGAGAGA 3652
Oy 421 CAGTGCAGAAATGCTATTTCGCGTGAACACACGCTTAA 456
Db 3653 CAGTGCAGAAATGCTATTTCGCGTGAACACACGCTTAA 3688

```

```

RESULT 2
LOCUS AE000205 10346 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
ACCESSION AE000205 U00096
VERSION AE000205.1 GI:1787265
KEYWORDS Escherichia coli K12
SOURCE Escherichia coli K12
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS 1 (bases 1 to 10346)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
TITLE JOURNAL
MEDLINE 97426617
PUBMED 9278503
REFERENCE
AUTHORS 2 (bases 1 to 10346)
Blattner, F.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 10346)
Blattner, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 10346)
Plunkett, G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

```

COMMENT

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: markborov@gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). ** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source

Location/Qualifiers
 1..10346
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 complement(<1..48)
 /note="IS3"

repeat_region

<1..5
 /note="efactor Sigma70; predicted +1 start at 1094689"

promoter

5..32
 /note="efactor Sigma70; predicted +1 start at 1094716"

promoter

69..392
 /gene="b1028"

gene

69..392
 /gene="b1028"

CDS

69..392
 /gene="b1028"
 /function="orf; Unknown"
 /note="0107; This 107 aa ORF is 27 pct identical (0 gaps) to 58 residues of an approx. 1888 aa protein FAS2_CANAL
 SW: P43098
 /codon_start=1
 /transl_table=11
 /product="orf; hypothetical protein"
 /protein_id="AAC74113.1"
 /db_xref="GI:1787266"
 /translation="MEDGVQAMRDYLAGLDIASDEHGVLMNTAKSEVAPSIKENLS
 LILHTFYKWTESLDPTPLNMPVAFLEISNKPVIGNMLNDPAVDQQRVHMCRAAFSD
 AKVKK"

CDS

275..299
 /gene="b1028"
 /note="efactor Sigma2; predicted +1 start at 1094983"

gene

389..1375
 /gene="ycdU"
 /note="synonym: b1029"

CDS

389..1375
 /gene="ycdU"
 /function="orf; Unknown"
 /note="0328; This 328 aa ORF is 19 pct identical (4 gaps) to 191 residues of an approx. 288 aa protein Y20_MYC8B
 SW: P47562"
 /codon_start=1
 /transl_table=11
 /product="orf; hypothetical protein"
 /protein_id="AAC74114.1"
 /db_xref="GI:1787267"

/translation="MIDYLPFRFODKRNLIYVAILGILGFWYKNAAGTTPSEDI
 GVSIGILALVYNFIPDLKAWAKCYTKNIDSWERKONKIEFLIOPVAFIS
 LMLSAMSWGLOLPSLALFLISLIGLVLFRMRTSYKQVIAVAKKYK
 SLTRYVLSVCISTVNLTLISPLRSDSVYBGMLTKSIIALLIGVLAFLF
 FLRSKRYAFLEOELIDFSSERNASTFFAKPLMLRFLILVIVMIVTIVSVL
 ATIVEMRIFBEAVFLCYVCLTYFFYCRFLMNDPMACDMYFRMGHPK"
 1253. .1280
 /gene="ycdW"
 /note="central position to predicted promoter: -104"
 /bound_moiety="Deor predicted site"
 1253. .1280
 /gene="ycdW"
 /note="No predicted promoter"
 /bound_moiety="Deor predicted site"
 1337. .1365
 /gene="ycdW"
 /note="factor Sigma70; predicted +1 start at 1096049"
 1494. .1745
 /gene="b1030"
 1494. .1745
 /gene="b1030"
 /function="orf; hypothetical protein"
 /note="orf; Unknown"
 /codon_start=1
 /product="orf, hypothetical protein"
 /protein_id="AAC74115.1"
 /db_xref="GI:1787268"
 /translation="MGSGPRCVGYTKKKVLSYELFFKYGEGSIDLSRSCGPPAH
 VHGSMLSPVDPBRSGLLIPVCAIHEKARTFVRL"
 1513. .1926
 /gene="ycdV"
 /note="synonym: b1031"
 1513. .1926
 /gene="ycdV"
 /function="putative structure; Not classified"
 /note="o137; This 137 aa ORF is 27 pct identical (1 gap)
 to 177 residues of an approx. 80 aa protein RS37_YEAST SW:
 P05759"
 /codon_start=1
 /transl_table=11
 /product="putative ribosomal protein"
 /protein_id="AAC74116.1"
 /db_xref="GI:1787269"
 /translation="MCNIRKSPYERTSSINAAVRGGLTRFAPAGSPULAFVCP
 GROLSTPGRGSPRCVQYKKKPVLSYELFFKYGEGSIDLSRSCGPPAHVHGL
 SMLSPVVEPRSGLLIPVCAIHEKARTFVRL"
 complement(2111. .2198)
 /gene="sexX"
 /note="synonym: b1032"
 complement(2111. .2198)
 /gene="sexX"
 /product="tRNA-Ser"
 /note="anticodon: GGA; CG Site No. 17878"
 /function="tRNA; tRNA"
 2269. .2296
 /note="factor Sigma70; predicted +1 start at 1096980"
 2338. .2365
 /note="factor Sigma70; predicted +1 start at 1097049"
 2355. .2383
 /note="factor Sigma70; predicted +1 start at 1097067"
 2393. .3370
 /gene="ycdW"
 /note="synonym: b1033"
 2393. .3370
 /gene="ycdW"
 /function="putative enzyme; Not classified"
 /note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
 to 177 residues of an approx. 312 aa protein YPRA_CORGL
 SW: P45637"
 /codon_start=1

/transl_table=11
 /product="putative dehydrogenase"
 /protein_id="AAC74117.1"
 /db_xref="GI:1787270"
 /translation="MRINPQRSQSMYDIIFFHPPTDQWTEALRKAIPQAVRAM
 KSGDSDADYALVWHPPEMLAGDKLKAVALGAVDYLISLKAHPMLNPSVPLR
 LEDTGMKQMEIVASVLMFRFDYRLOQNSHMOPLPEYHREDPTIGICAVL
 GSKVAQSLQTMFPLRCMSRTKRSWGVOSFAGRELSAFISCCVLTILNLEPTEV
 GLNQULBEKLPDGAVIDNLARGVHVEDDLAALDSGVKAMLDVNERPLPESP
 LMQHPRTIIPVVAITRPAAVEYISRTIAQLEKERRCGVDRARG"
 3278. .3308
 /gene="ycdW"
 /note="factor Sigma70; predicted +1 start at 1097992"
 3383. .3410
 /note="factor Sigma70; predicted +1 start at 1098094"
 3396. .3423
 promoter

Query Match 100.0%; Score 456; DB 1; Length 10346;
 Best local similarity 100.0%; Pred. No. 2.7e-116; Mismatches 0; Indels 0; Gaps 0;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTGTTATTTATGATTTACATATCTGGCTGGGATTCAGCC 60
 Db 8497 ATGAAAAACAATTTGTTATTTATGATTTACATATCTGGCTGGGATTCAGCC 60
 QY 61 GCAGCGGTTATGATTTAGCTAATTCAGATATTAATTCGGGTAATGATTAAGTAA 120
 Db 8557 GCAGCGGTTATGATTTAGCTAATTCAGATATTAATTCGGGTAATGATTAAG 120
 QY 121 TCTTCATTTATCAGGCGCATTAATGCTACAGCTGGACATAATATAGTCTCAGTTA 180
 Db 8617 TCTTCATTTATCAGGCGCATTAATGCTACAGCTGGACATAATATAGTCTCAGTTA 180
 QY 181 CGGAGGAGGCTCAAACTTTGGCGTGTGGCGAAGAGTGTGCAACCGGCGCA 240
 Db 8677 CGGAGGAGGCTCAAACTTTGGCGTGTGGCGAAGAGTGTGCAACCGGCGCA 240
 QY 241 AAGATTGACGACAGAGATTTAATCTTGCTATATATGATGAGCGGCGAGTCCAA 300
 Db 8737 AAGATTGACGACAGAGATTTAATCTTGCTATATATGATGAGCGGCGAGTCCAA 300
 QY 301 GATGCGATTTTTCGCAAGTGTGTTATGTAATCTGATGATTAATCAGAAAGTTCT 360
 Db 8797 GATGCGATTTTTCGCAAGTGTGTTATGTAATCTGATGATTAATCAGAAAGTTCT 360
 QY 361 GGTATTAAGCAATATTTACAGATGATGCTCAAAAAAGCGCAATTTGATGACAGA 420
 Db 8857 GGTATTAAGCAATATTTACAGATGATGCTCAAAAAAGCGCAATTTGATGACAGA 420
 QY 421 CAGTGGCAATGCTTATTCGCTGATGACACAGCTTAA 456
 Db 8917 CAGTGGCAATGCTTATTCGCTGATGACACAGCTTAA 8952

RESULT 3

D90741 15047 bp DNA linear BCT 25-DBC-2002
 LOCUS Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
 DEFINITION D90741 AB001340
 ACCESSION D90741.1 GI:1651509
 VERSION Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
 csgA; ycdB; cIs; nov; mdcG.
 KEYWORDS Escherichia coli K12
 SOURCE Escherichia coli K12
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE

AUTHORS Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
 Ikeno, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
 Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
 Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
 Nishio, Y., Saito, N., Sampei, G., Seki, Y., Teganu, H., Takemoto, K.,
 Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T.

TITLE A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map

JOURNAL DNA Res. 3 (3), 137-155 (1996)

MEBLINE 97061202

PUBMED 8905232

REFERENCE 2

AUTHORS Ikema, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kaibuchi, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Mikita, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.

TITLE The systematic sequencing of the *Escherichia coli* genome in Japan

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 15047)

AUTHORS Mori, H.

JOURNAL Direct Submission

REFERENCE Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info., 8916-5 Takayama, Ikoma, Nara 630-01, Japan

(E-mail: hmori@gc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)

COMMENT Collaboration Information:

Project:

The Japan E. coli genome DNA sequencing project

Group:

The Japan E. coli genome DNA sequencing group

Members: (1995.4 - 1996.3)

Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasai, H., Kaibuchi, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Mikita, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.

Headed by:

Name: Takeshi Horiuchi

Address: National Institute of Basic Biology, Okazaki, 444, Japan

E-mail: kishior@nibb.ac.jp

Information operator:

Name: Hirotsada Mori

Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan

E-mail: hmori@gc.aist-nara.ac.jp

URL:

The Japan E. coli genome database

http://bwa3.aist-nara.ac.jp

Location/Qualifiers

1. 15047

/organism="Escherichia coli K12"

/mol_type="genomic DNA"

/strain="K12"

/db_xref="taxon:8333"

/map="23.7-24.0 min"

/clone="Kohara clone #231"

/note="Nucleotide position 1097101-1112147 from the initiation site of Theta (0 min.). This clone is from Kohara lambda miniset library."

<1. 276

/notes="unnamed protein product; ORF_ID:0230#5"

/codon_start=1

/transl_table=1

/protein_id="BAA35822.1"

/db_xref="GI:4062605"

/translation="DLFSSSEALSTFAKPLMLRLFILVIEVMMITLVSLATLVE WRWFEAYFLICVPCILYFPCRFPLHNDPMMACDMYFRMGHPNK"

1152. 2090

/note="ORF_ID:0231#1

similar to SwissProt Accession Number P45637"

/codon_start=1

/transl_table=1

CDS

gene

/product="Hypothetical 33.0 kd protein in prob-proa intergenic region."

/protein_id="BAA35823.1"

/db_xref="GI:4062606"

/translation="MDIIFHPTPTQWMEALRKALIPQARVANKSGDNDADYALV WHPEVMEAGDLKAVFALGAGVDSILSKLOAHPEMLPSVPLFLEPTGMEQOEY AVSOVILWFRFRFDYRIQNSSHOPEYREDEPTGILAGVSKVSAQSLQWRF PLRWSRPRKSPROCVSPAGREELSAFLSQCRVLINLPTNPETGTGINOULECPD GAYILNLRGVHVEDDLALDSDGRGAMLDVFNREPLPESPLMOHPRTTTPHV AATRPAAVYISRTTIAOLEKGRVCGQVDPRAGY"

2145. 2882

/note="unnamed protein product; ORF_ID:0231#2"

/codon_start=1

/transl_table=1

/protein_id="BAA35824.1"

/db_xref="GI:4062607"

/translation="MYPVLDLMTVASTHAYSTLSDYIAQAKQKIKLPAITDHPDM EDAPHMHFIMNRIMPRVDSVILRGIEANIKNVDSIDCSKMFSLDLIIAGFHE PVAPHDKATNTQMIATIASGVHIIHPNPYKEIDVKAVAAAGHOVALEINIS SPISRRGSENDCEVAAVAPDAGGVALGSDSHTAFTMGSEFBECLKILDVDFPPER ILNVSPRLNLFLESKGMATPFAEFDL"

2906. 3460

/note="ORF_ID:0231#3

similar to PIR Accession Number F64035"

/codon_start=1

/transl_table=1

/product="Hypothetical protein H11543"

/protein_id="BAA35825.1"

/db_xref="GI:4062608"

/translation="MNEFSILCRVLGSLYRPODPLVPLFTLIREGLAANPLQ DELTTRLOKSCDMQVADYALPIGECAPPRSAWBEATATAVAPAFSEGMPL ADPRAFLCTKLLASWLEDOSTDESBALTELSVILLPCGAFLEKVEAHATTFFV RTMAPLTRDAISAMDELBEDSE"

3514. 4053

/note="unnamed protein product; ORF_ID:0231#4"

/codon_start=1

/transl_table=1

/protein_id="BAA35826.1"

/db_xref="GI:4062609"

/translation="WCSISFMAAFSAIRGMNILLSTAITTGLISGWAVASLGLL SWAGFLCTAFACPDGLKGLISAATLGSVMAWVILYGSALAPLEITLGYVITG IVAPFLMCIOAKOLLISFVPGTFIGACATFAGQGDWKLVPALGLLFGVAMKSGLM LAARSATFAHREDEIKNKA"

complement(4117. 4950)

/gene="cs9c"

complement(4117. 4950)

/gene="cs9c"

/note="ORF_ID:0231#5

similar to SwissProt Accession Number P52103"

/codon_start=1

/transl_table=1

/product="Curli production assembly/transport component CsgG precursor."

/protein_id="BAA35827.1"

/db_xref="GI:4062610"

/translation="MORFLIVAVMLTSGCLTAPKRAARPLMPRAQSYKDLTLPA PTGKITVSVINIDETSCQFKPYPSANSTAVPGSATMLTALDSKRFITLBEQGLQ NLNLERKIRIAQENGTVAINNRIPLOSLTAANTVGSIIIGYSNVKSGGVARFYG IGADTOYLOQIDAVNLVNVNSTEILSVTSKTIISYEQADVFPERIDVORLBGR VGYTSNBPVMLCLMSALETGIVIFLINDIGRGLDNLKAKERQNDILVKYHMSVPP S"

complement(4977. 5393)

/gene="csgF"

complement(4977. 5393)

/gene="csgF"

/note="ORF_ID:0231#6

similar to SwissProt Accession Number P52104"

/codon_start=1

/transl_table=1

/product="Curli production assembly/transport component CsgF precursor."

/protein_id="BAA35828.1"

/db_xref="GI:4062611"

CDS

CDS

FEATURES

SOURCE

COMMENT

JOURNAL

REFERENCE

TITLE

AUTHORS

REFERENCE

JOURNAL

MEBLINE

PUBMED

AUTHORS

TITLE

JOURNAL

REFERENCE

JOURNAL

gene
CDS
/translation="MRVKAHVLLMLISPLSWAGTMTPOFRNPEGANNGAFLINS
AOQNSYKDSYNDPFGIETPSALNDPFOIIOIIGLLSININGKPGRMVNDYIV
DIANDGQLOLVNDRKGTSTIOVGLNNSTDP"
complement(5418..5807)
/gene="csge"
complement(5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsgB precursor."
/protein_id="BA035829.1"
/db_xref="GI:4062612"
/translation="MKRYRRLVVAEFLPAANGHAEVEVGLITDHTVSSIGHDFY
RAPSDSDSDYGNLTINRPSRWSWITTVQDVLPQFLPLRDEKTVVRL
IOBEALNRQINQMLSTGDLAHDPE"
complement(5812..6462)
/gene="csge"
complement(5812..6462)
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=11
/product="Probable csgB operon transcriptional regulatory
protein."
/protein_id="BA035830.1"
/db_xref="GI:4062613"
/translation="MPEVHSIHGHTLLITSSLOANTLLQHLKOSLAITGKLNIG
RSLDDISSGIIILDMEADKLIHWMOTLSRKNNIKILANTPENYEDENP
HINQVFSMEDORRYVNGLOGVKRGCFYTKLASYLTHSGNRYSTESALLTHRE
KEILNKLRIGASNNELARSLFSENTVKTHLVLPKIAVKNRQVSMADNLRP"
complement(7672..7672)
/gene="csge"
complement(7672..7672)

gene
CDS
Query Match
Best Local Similarity 100.0%; Score 456; DB 1; Length 15047;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAAAACAATTTATTTATGATTTAAACATATCTGGCGCCCTGGATTGACCC 60
7217 ATGAAAAACAATTTATTTATGATTTAAACATATCTGGCGCCCTGGATTGACCC 60
61 GCAGCAGTTATGATTTAGCTAATTCAGATATACTCGCGTAAATGATTGAGTAAG 120
7277 GCAGCAGTTATGATTTAGCTAATTCAGATATACTCGCGTAAATGATTGAGTAAG 120
121 TCTTCATTTATCAGGACCAATATTTGCTCAAGCTGGACTAATAAGTGTCAAGTTA 180
7337 TCTTCATTTATCAGGACCAATATTTGCTCAAGCTGGACTAATAAGTGTCAAGTTA 180
181 CGGCGAGGAGCTCAAACTTTGGCGTTGGCGCAAGAGTATGCAACCGGGCA 240
7397 CGGCGAGGAGCTCAAACTTTGGCGTTGGCGCAAGAGTATGCAACCGGGCA 240
241 AAGATTGACGACAGAGATATTAACCTTGCAATATTTGATCAGCGCGGAGTCCAC 300
7457 AAGATTGACGACAGAGATATTAACCTTGCAATATTTGATCAGCGCGGAGTCCAC 300
301 GATGCCGATATTTGCAAGGTGCTTATGTTAATCTGGGATATTTATCCAGAAAGTTCT 360
7517 GATGCCGATATTTGCAAGGTGCTTATGTTAATCTGGGATATTTATCCAGAAAGTTCT 360
361 GGTATTAAGCAAAATTTATCAACGATGTGTTACTCAAAAAACGCAATTTGATGTCAGAGA 420
7577 GGTATTAAGCAAAATTTATCAACGATGTGTTACTCAAAAAACGCAATTTGATGTCAGAGA 420
421 CAGTGCAGATGCTATTTGCGGTGACACAGCTTAA 456
7637 CAGTGCAGATGCTATTTGCGGTGACACAGCTTAA 456
7672 CAGTGCAGATGCTATTTGCGGTGACACAGCTTAA 456

RESULT 4
AF275733
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
Ulrich, G.A., Keen, J.E. and Elder, R.O.
Mutations in the csgD promoter associated with variations in curli
expression in certain strains of Escherichia coli O157:H7
Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
AF275733.1 GI:14039399

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
2 (bases 1 to 1711)
Ulrich, G.A., Keen, J.E. and Elder, R.O.
Direct Submission
Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
Research Center, State Spur 18D, Clay Center, NE 68933, USA
1..1711
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
757..1212
/gene="csgB"
757..1212
/gene="csgB"
/codon_start=1
/transl_table=11
/product="CsgB protein"
/protein_id="AAK33211.1"
/db_xref="GI:14039400"
/translation="MKKLLFMWLTILGAPGIAAAGVGLANSENPANVELSGSNN
QAATIGAGTNNASQIRQGSKLAVVDESSNRAKIDQGDYNAITIDQGSANDA
SISQYAGNTAMIIQKSGSNKANTIVYGTQATVAVRQSMALRVQR"
1253..1711
/gene="csgA"
1253..1711
/gene="csgA"
/codon_start=1
/transl_table=11
/product="CsgA protein"
/protein_id="AAK33212.1"
/db_xref="GI:14039401"
/translation="MKLVKVAATAATVYFSSALAGVFPVYGGGNGHGGGNNSPNS
ELNIYVGGNSALATQADARNSDITTTGSGGNGADVQGSDDSDSIDTQGFNSA
TLQDNGDSIMTYKQVGGNGAIVDOTANSNTVNTVQGFNNATAHQY"

gene
CDS

ORIGIN

Query Match
Best Local Similarity 99.6%; Score 454.4; DB 1; Length 1711;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGAAAAACAATTTATTTATGATTTAAACATATCTGGCGCCCTGGATTGACCC 60
757 ATGAAAAACAATTTATTTATGATTTAAACATATCTGGCGCCCTGGATTGACCC 60
61 GCAGCAGTTATGATTTAGCTAATTCAGATATACTCGCGTAAATGATTGAGTAAG 120
817 GCAGCAGTTATGATTTAGCTAATTCAGATATACTCGCGTAAATGATTGAGTAAG 120
121 TCTTCATTTATCAGGACCAATATTTGCTCAAGCTGGACTAATAAGTGTCAAGTTA 180

Db	877	CTTCATTTAATCAGGCAAGCCATAATGTTCAAGCTGGGACCTAAATATATAGTCTCAGTTA	936
Qy	181	CGGCAGGAGAGGCTCAAAACTTTTGGCGGTTGTTGGCAAGAAAGTAGTACACCCGGCA	240
Db	937	CGGCAGGGAGGCTCAAAACTTTTGGCGGTTGTTGGCAAGAAAGTAGTACACCCGGCA	996
Qy	241	AAGATTGACCAAGACAGAGATTATTAACCTTGATATATTGATCAGGCGGGCAGTCCCAAC	300
Db	997	AAGATTGACCAAGACAGAGATTATTAACCTTGATATATTGATCAGGCGGGCAGTCCCAAT	1055
Qy	301	GATGCCAGTATTTGCCAAGGTGCTTATGCTAATACTCGCAGTATTAATCCAGAAAGTCTT	360
Db	1057	GATGCCAGTATTTGCCAAGGTGCTTATGCTAATACTCGCAGTATTAATCCAGAAAGTCTT	1114
Qy	361	GGTATAAAGCAAAATATTACACAGTATGTACTCAAAAAACGGCAATTGTAGTCAGAGA	420
Db	1117	GGTATAAAGCAAAATATTACACAGTATGTACTCAAAAAACGGCAATTGTAGTCAGAGA	1176
Qy	421	CAGTCGCAAAATGGCGTATTCGGCTGACACAAGCTTAA	456
Db	1177	CAGTCGCAAAATGGCGTATTCGGCTGACACAAGCTTAA	1212

RESULT 5	AE005315	10190 bp	DNA	linear	ECT 21-MAR-2001
LOCUS	AE005315				
DEFINITION	AE005315	10190 bp	DNA	linear	ECT 21-MAR-2001
	Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139 of 155.				
ACCESSION	AE005315	AE005174			
VERSION	AE005315.1	GI:12514572			
KEYWORDS					
SOURCE	Escherichia coli O157:H7 EDL933				
ORGANISM	Escherichia coli O157:H7 EDL933				

REFERENCE
AUTHORS
1 (bases 1 to 10190)
Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,

TITLE Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7
JOURNAL *Nature* 409 (6819), 529-533 (2001)

REFERENCE 2 (bases 1 to 10190)
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,

TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

```

source
1. .10190
location=genbank
organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohemorrhagic"
617. .1072
/gene="ecgB"
617. .1072
CDS
617. .1072
/note="synonym: Z1675"
617. .1072

```

```
/function="structural component; Cell exterior  
constituents: Surface structures"  
/note="Residues 1 to 151 of 151 are 100.00 pct identical  
to residues 1 to 151 of 151 from Escherichia coli K-12
```

```

Strain MG1655: B1041"
/codon_start=1
/transl_table=11
/product="minor curlin subunit precursor, similar to CsgA"
/protein_id="AAG55787.1"
/db_xref="GI:12514573"
/translation="MKKKLLPMLTITIGAPGIAAAGYDIANSEYFAVELSKSSFN
QAIIQAGTNNASQLRQGGSKLAVVAOESSNRATIDGTGYNALYIDQASANDA
SISQAYGNTAMITLQKSGNKANITQGTOKTAIVVQRSGMAIRVQR"
1113..1571
/gene="csgA"
/note="synonym: Z1676"
1113..1571
CDS

```

```

/function="regulator; Cell exterior constituents: Surface
structures"
/note="residues 1 to 152 of 152 are 96.71 pct identical to
residues 1 to 151 of 151 from Escherichia coli K-12 Strain
MG1655. B1042"
/codon_start=1
/crystal_table=1
/product="curlin major subunit, coiled surface structures;
cryptic"
/protein_id="MAG55788.1"
/db_xref="gi:12514574"
/translation="MLKLKAAIAIVFSSGALAGVPPQYGGGGGNGHGGGNGSPNS
ELNIYVGGGNSALALQADARNSDLTITHGGGNGAGDVGGGSDSSIDLIRQFSGSA
TLDDWNGDSDMTKQFGGGNGGAADVDTAASNTVNTQVGFNNATAHQ"
1630. 1962
/gene="csgC"
/note="synonym: Z1677"
1630. 1962

```

```

Gene      2083. .2394
          /gene="Z1678"
          2083. .2394
          /gene="Z1678"
          /function="orf; Unknown function"
          /note="Residues 1 to 103 of 103 are 98.05 pct identical to
          residues 1 to 103 of 103 from Escherichia coli K-12 Strain
          MG1655: B1044"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="PAG55790.1"
          /db_xref="GI:12514576"
          /translation="MPRPFLNLSMLGSLFFPFPIAGSTAGCGVHHFYGOIVEPACDV
          STGSPVEMNCPQNGSVPGKTYSSKALMSGNVAQAIASVKVQYLDKQKKLVAMNIEY
          N"
Gene      2489. .3022
          /gene="Z1679"
          2489. .3022
          /gene="Z1679"
          /function="orf; Unknown function"
          /note="Residues 1 to 177 of 177 are 100.00 pct identical
          to residues 1 to 177 of 177 from Escherichia coli K-12
          Strain MG1655: B1045"
          /codon_start=1
          /transl_table=11
          /product="putative polyprotein"
          /protein_id="PAG55791.1"

```

gene
CDS
/db_xref="GI:12514577"
/translation="MKTRIHVQGDITKLAADVIVNANPSIMGGCGDGAIRHRAAP
ALPDACLVKQOQDCEPTGHAIVITLADLPAKAVITTPGVRGSGEDDOLQAVL
NSLRIVANSTSVAFPAISGVYGPRAAAAIIVKIVSEPTIRHALPEQYFVCYD
2964. .4445
/gene="ymcH"
/note="synonym: Z1680"
2964. 4445
/gene="ymcH"
/function="putative enzyme; Not classified"
/note="Residues 1 to 493 of 493 are 99.59 pct identical to
MG1655: B1046"
/codon_start=1
/transl_table=11
/product="putative synthase"
/protein_id="AAG55792.1"
/db_xref="GI:12514578"
/translation="MKKTPTSDKSLPNKEMNDPLASAVPLCSQHPGQGLFPL
EKSLDAFARYLAEMSEHTLDVOYIMODMSGRLLPSALLAARGVRRLDLDN
NTPGLDILRLDHPRIEVLNPPSRLLRPLGYITDSLRNRNKKSFIVGDV
TLVGRNIGDAVFGAGEPLPSDIDVAHIGVEDVADDFARYCKSVPLQOYVDV
PEGEMADRIELPASWHDNMTARYLRMSESPFINLVDGTLPLIAKTRILSDPAK
GEGKAKRHSILPQRLFDIMGSESRIDIISSYFVPRAGVAOLLRVRRKXVATITN
SLAANDVAVVAGTARWKRLKILRYGVELKPTROSSLHDHRTGNSGSLAKT
PSIDKTVFISFNFDPSRLNTEMGFVIESETLAQILDIDKIQSQYDAAWQLRDLR
MGRINWDRHAKKEIVLKSEFATSPFMRVWVRLASILPEVMI"
complement (4453. .5610)
/gene="ymcD"
/note="synonym: Z1681"
complement (4453. .5610)
/gene="ymcD"
/function="orf; Unknown function"
/note="Residues 1 to 385 of 385 are 99.22 pct identical to
MG1655: B1047"
/residues 1 to 385 of 385 from Escherichia coli K-12 Strain
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55793.1"
/db_xref="GI:12514579"
/translation="MNPVPAQREYFXDSIRAMLMILGIPHISLISYHTWYNSAP
SLMTIRNDPIHSEFMQVFEVLSGYFSLMLFLARYPLKMKKVVVERVGIPLTAIFL
TLPQITLQYVKGAKESWPGSLYDKYNTLAMELISHLMPLVLYVNTTLCWITFRI
RNKLNSDKTNKFSMVKLSYFLCTIGIGVAVIRRTFITYPILSNGTENFTVQTL
FYLPEFIIIGALAFIPHLKALFTTPSGCTAALAAVAVLNRGSGDAMMYETES
VITVNLGLMNVVVSFGHRLNFOASRVYFVNASLFTLVHPLTLFPGAVTPIHI
6003. 7538
/gene="ymcG"
/note="synonym: Z1683"
6003. 7538
/gene="ymcG"
/function="enzyme; Osmotic adaptation"
/note="Residues 1 to 511 of 511 are 99.80 pct identical to
MG1655: B1048"
/residues 1 to 511 of 511 from Escherichia coli K-12 Strain
/codon_start=1
/transl_table=11
/product="periplasmic glucans biosynthesis protein"
/protein_id="AAG55794.1"
/db_xref="GI:12514580"
/translation="MKMKRMLSAAVMTLITSSSNAFISIDVAKQASLAGKVEAPK
SNLPSYRDMKADYQOIOFNHDKYNNMLKTPLEFVHQMGEDPPVNIINVTATA
GAGQVYGLSARGAIDTALPSGEFPRKEFWIRPXTDKRLITVIALDPSRATGAY
KTFVWGPRTVDVOSKTYLNDKVGKLGVALTMTPLFGPQPSPPANNVYPRHDSNG
LSIHNGSGMIRPLNPKHLAVSFSMENQGGGLORGRDPSREDDDRIDRDS
ANVTPEKGSGVELVEIPNDENDNIVAVYTPDLPGRGKNAFKTITTSRSD
KHAADNMAVQOTRSTGDKOSNLIROPOTTAFTVDFGAKMKCLPDTPTVATGS
IGDNCEIVESTVRVNPYTKGRLVWRKIVADAKTTTMRALVNAODLTSETWSYOLP
ANE"

gene
CDS
7531. .10074
/gene="mdcH"
/note="synonym: Z1684"
7531. .10074
/gene="mdcH"
/function="enzyme; Osmotic adaptation"

Query Match 99.6%; Score 454.4; DB 1; Length 10190;

Best Local Similarity 99.8%; Pred. No. 7,4e-116;

Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	ATGAAAAACAATGTTATTATGATGATTAACATATCTGGCGCTGGAGTTCAGCC	60
Db	617	ATGAAAAACAATGTTATTATGATGATTAACATATCTGGCGCTGGAGTTCAGCC	676
Qy	61	GCACGAGTTATGATTAATGATTAATTCAGATTAATTCGCGTAAATGAATGATGATAG	120
Db	677	GCACGAGTTATGATTAATGATTAATTCAGATTAATTCGCGTAAATGAATGATGATAG	736
Qy	121	TCTTCATTTATGACGACGACCAATTAATGATTAATTCGCGTAAATGAATGATGATAG	180
Db	737	TCTTCATTTATGACGACGACCAATTAATGATTAATTCGCGTAAATGAATGATGATAG	796
Qy	181	CGCAGAGGAGGCTCAAAATCTTTGGCGGTGTTGGCAAGAGGTAAGCAACGGGCA	240
Db	797	CGCAGAGGAGGCTCAAAATCTTTGGCGGTGTTGGCAAGAGGTAAGCAACGGGCA	856
Qy	241	AGATTGACACAGACGAGATTAATTAATTCATATTAATGATGATGACGCGGCACTGCCA	300
Db	857	AGATTGACACAGACGAGATTAATTAATTCATATTAATGATGATGACGCGGCACTGCCA	916
Qy	301	GATGCCAGATTTGGCAAGGCTTATGATTAATTCGATGATTAATTCGAGAAAGTTCT	360
Db	917	GATGCCAGATTTGGCAAGGCTTATGATTAATTCGATGATTAATTCGAGAAAGTTCT	976
Qy	361	GATTAATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	420
Db	977	GATTAATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1036
Qy	421	CAGTCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	486
Db	1037	CAGTCGCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	1072

gene
CDS
/db_xref="GI:12514577"
/translation="MKTRIHVQGDITKLAADVIVNANPSIMGGCGDGAIRHRAAP
ALPDACLVKQOQDCEPTGHAIVITLADLPAKAVITTPGVRGSGEDDOLQAVL
NSLRIVANSTSVAFPAISGVYGPRAAAAIIVKIVSEPTIRHALPEQYFVCYD
2964. .4445
/gene="ymcH"
/note="synonym: Z1680"
2964. 4445
/gene="ymcH"
/function="putative enzyme; Not classified"
/note="Residues 1 to 493 of 493 are 99.59 pct identical to
MG1655: B1046"
/codon_start=1
/transl_table=11
/product="putative synthase"
/protein_id="AAG55792.1"
/db_xref="GI:12514578"
/translation="MKKTPTSDKSLPNKEMNDPLASAVPLCSQHPGQGLFPL
EKSLDAFARYLAEMSEHTLDVOYIMODMSGRLLPSALLAARGVRRLDLDN
NTPGLDILRLDHPRIEVLNPPSRLLRPLGYITDSLRNRNKKSFIVGDV
TLVGRNIGDAVFGAGEPLPSDIDVAHIGVEDVADDFARYCKSVPLQOYVDV
PEGEMADRIELPASWHDNMTARYLRMSESPFINLVDGTLPLIAKTRILSDPAK
GEGKAKRHSILPQRLFDIMGSESRIDIISSYFVPRAGVAOLLRVRRKXVATITN
SLAANDVAVVAGTARWKRLKILRYGVELKPTROSSLHDHRTGNSGSLAKT
PSIDKTVFISFNFDPSRLNTEMGFVIESETLAQILDIDKIQSQYDAAWQLRDLR
MGRINWDRHAKKEIVLKSEFATSPFMRVWVRLASILPEVMI"
complement (4453. .5610)
/gene="ymcD"
/note="synonym: Z1681"
complement (4453. .5610)
/gene="ymcD"
/function="orf; Unknown function"
/note="Residues 1 to 385 of 385 are 99.22 pct identical to
MG1655: B1047"
/residues 1 to 385 of 385 from Escherichia coli K-12 Strain
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55793.1"
/db_xref="GI:12514579"
/translation="MNPVPAQREYFXDSIRAMLMILGIPHISLISYHTWYNSAP
SLMTIRNDPIHSEFMQVFEVLSGYFSLMLFLARYPLKMKKVVVERVGIPLTAIFL
TLPQITLQYVKGAKESWPGSLYDKYNTLAMELISHLMPLVLYVNTTLCWITFRI
RNKLNSDKTNKFSMVKLSYFLCTIGIGVAVIRRTFITYPILSNGTENFTVQTL
FYLPEFIIIGALAFIPHLKALFTTPSGCTAALAAVAVLNRGSGDAMMYETES
VITVNLGLMNVVVSFGHRLNFOASRVYFVNASLFTLVHPLTLFPGAVTPIHI
6003. 7538
/gene="ymcG"
/note="synonym: Z1683"
6003. 7538
/gene="ymcG"
/function="enzyme; Osmotic adaptation"
/note="Residues 1 to 511 of 511 are 99.80 pct identical to
MG1655: B1048"
/residues 1 to 511 of 511 from Escherichia coli K-12 Strain
/codon_start=1
/transl_table=11
/product="periplasmic glucans biosynthesis protein"
/protein_id="AAG55794.1"
/db_xref="GI:12514580"
/translation="MKMKRMLSAAVMTLITSSSNAFISIDVAKQASLAGKVEAPK
SNLPSYRDMKADYQOIOFNHDKYNNMLKTPLEFVHQMGEDPPVNIINVTATA
GAGQVYGLSARGAIDTALPSGEFPRKEFWIRPXTDKRLITVIALDPSRATGAY
KTFVWGPRTVDVOSKTYLNDKVGKLGVALTMTPLFGPQPSPPANNVYPRHDSNG
LSIHNGSGMIRPLNPKHLAVSFSMENQGGGLORGRDPSREDDDRIDRDS
ANVTPEKGSGVELVEIPNDENDNIVAVYTPDLPGRGKNAFKTITTSRSD
KHAADNMAVQOTRSTGDKOSNLIROPOTTAFTVDFGAKMKCLPDTPTVATGS
IGDNCEIVESTVRVNPYTKGRLVWRKIVADAKTTTMRALVNAODLTSETWSYOLP
ANE"

RESULT 6
AP002554
LOCUS
DEFINITION
Escherichia coli O157:H7 DNA, complete genome, section 5/20.
VERSION
AP002554.1 GI:13360491
KEYWORDS
SOURCE
ORGANISM
Escherichia coli O157:H7
Escherichia coli O157:H7
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1
Makino, K., Yokoyama, K., Kubota, Y., Yuesudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tetsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C. and Shinagawa, H.
Complete nucleotide sequence of the prophage ϕ 2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE
20198780
PUBMED
10734605
REFERENCES
Ohnishi, M., Murata, T., Nakayama, K., Kubota, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655

JOURNAL MEDLINE PUBMED REFERENCES AUTHORS	Syst. Appl. Microbiol. 23 (3), 315-324 (2000) 20557356 11108008 3 Yokoyama, K., Makino, K., Kubota, Y., Matanabe, M., Kimura, S., Yusuf, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.
TITLE	Complete nucleotide sequence of the prophage VPI-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 strain derived from the Sakai outbreak
JOURNAL MEDLINE PUBMED REFERENCES AUTHORS	Gene 258 (1-2), 127-139 (2000) 20564182 1111050 4 Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
TITLE	Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL MEDLINE PUBMED REFERENCES AUTHORS	DNA Res. 8 (1), 11-22 (2001) 21156231 11258796 5 (bases 1 to 327773) Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp), Tel: 81-6-6879-8365, fax: 81-6-6879-2047)
COMMENT FEATURES SOURCE	genome project. Location/Qualifiers 1..327773 /organism="Escherichia coli O157:H7" /mol_type="genomic DNA" /strain="O157:H7" /sub_strain="RIMD 0509952" /db_xref="taxon:83334" 150..2258 /gene="EC61032" 150..2258 /note="similar to YCBY ECOLI gi 1787181 percent identity 99 in 702 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=1 /product="putative oxidoreductase" /protein_id="BAB34455.1" /db_xref="GI:13360492" /translation="MNSLFASTARGDELKTELLENLGAVERCOVVGSGHFKGPNRLV YQSIWMSRLASRIWLPGECKVYSDLDLYLQVAINTEMNPGATPRVHFGSLNDTV RNSQYGAWKVDATVDAFTKRLPREPNDRDAPDIRVWVWLKSTASIALDLSGDHL LRGRDRAGAPIKETTLAALVMSGWPGRPLDPMCGSTLLIEAMLTADAPGL HRRGWSGSGMAIDHEALWQEVKASQTRARGLAESHSFYGSDSDARVLOTARNGP LAGICELTFEVKVAULTNPLPKPEYGTVSNPPYGERLDSBPALILHSLRIRIK NPGGAMNLSFSAIPDLSCLQLRADQYKAKNGPLDCVQNGVVAESTPDSKRAMA EDVANRLRKTKREKAROGICRYLVDADLERYVAVNDRYADWVYVQYAPPKTI DAKAKRRLPDTIAATISVGLIAPNKVLTKRKQKNGQYQKLGKKEFLVETVNA HLWNLVDYLDLGLPLDRIARMLQMSKDFLNLFSYGSATVHAGLGASSTT VDMRSRYLWAEARNRLNLGTLGRARHLQADCLMLREANQFDLIFDPTFENSKR MEDAFVQRDHIALMKDLKRLIRAGTIFSNNRGFRMDGLAKLGLKQETQKT LGGDFARROIHNCILITAA"
gene	2270..4177 /gene="EC61033" 2270..4177 /note="similar to UUP ECOLI gi 1787182 percent identity 99 in 635 aa (Conserved in E.coli K-12)"
CDS	
gene	4307..5560 /gene="EC61034" 4307..5560 /note="similar to POIA ECOLI gi 1787183 percent identity 99 in 417 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=1 /product="paraguat-inducible protein A" /protein_id="BAB34457.1" /db_xref="GI:13360494" /translation="WCEHHDAKIIICSCQDMVLALPRLHEGQRAACRCCITLVAM DAPQRPAYALALFMLILSNLPFPVNMVAVGTSITLLEIGVASEDYASLGT FLFVOLVPAFCILITILLVNRALPRLEKQRLAVFLQKTMGAELFLAGVVSFV KLAVGSIIGVSSFLPWLFCVQLQRAFCQVDRMLMDIDAPMELQPLRGVIGIR QGLRSCGCTAIIIPADEPVCRCCTKGVRRRNLQMTLALVLSIMLYLPALPIIM VTDLGKMSPTIIAGVILIMSSESVAAVIFASIMVPTLMKIAIIMLCMDAKHG KRDSERMLIYEVAFYGRMSMIDVPFIAYLSALVRMGIMSTIPANGALFALVIM TFEAMTFDEPLSDRQPSSEHES"
gene	5565..7205 /gene="EC61035" 5565..7205 /note="similar to PQIB ECOLI gi 1787184 percent identity 99 in 546 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=1 /product="paraguat-inducible protein B" /protein_id="BAB34458.1" /db_xref="GI:13360495" /translation="MESNNGEAKIOKVMNSPWIPIVYALIGAWLYFYHSHQGE VTLITAAIEGKRTYIKRSVVDVGVESATLADLTHTHEIKRLNSGMEKLAKHT VFWVVKPQIGREGISGLTLLSGVILELOGAGSKKDKYDLDSPLAPDAGIRV VLDSKRGQISRPDPVPRGVRSVSTSPFDQKRIYSQVLFNADYDVLVSNRP WDSGLAVDITLSAKRVKMSGLTLLSGSVFVPEBLSLQAPAPPTAFVLYDDQS IQSLYLDYDLMLEFFDQSVRGLQGAPEFRGIRLQTVSKVPFAFNMQTFDDYR IPVILIRIEPRNLQLEBNADVHLELKRGLRGLKTNGLVTGALVYDLDYDPPT PATIGLIEFNGYQIIPVSGLAQIIOQLMEALDKINKLPLPMIEQATSTLSGORT MKQVLTQPLTKLMEASNLVFPKAKKQDPEPRGAG"
gene	7217..7765 /gene="EC61036" 7217..7765 /note="similar to YWBA ECOLI gi 1787185 percent identity 98 in 182 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=1 /product="hypothetical protein" /protein_id="BAB34459.1" /db_xref="GI:13360496"
CDS	

	gene	/translation="MTIALMLAGSSGEINKNNYQLPVQSGTGSTASOGNRLITVTS OAVPDIYLANGVTYTQISDVKVIANNINIMASPLDOOLRNTVALIETQLGSMIVYAGC PLCSADPTIANVTVLEFNRYDGKRVIVSGEWMILNHQGQLIKRPFLRESIQVDGTIDEMV KULAGWSEDAASAISOIEKRLP" 8021..8188 /gene="ECs1037" 8021..8188 /gene="ECs1037" /note="identical to RMF_ECOLI gi 1787186 (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=11 /product="ribosome modulation factor" /protein_id="BAB34460.1" /db_xref="GI:13360497" /translation="MKRKRDRIERAHQRGVQNGIAGRSKMPCYDTLNQRSQWLGM REAAADVYMA" complement(8258..8776) /gene="ECs1038" complement(8258..8776) /gene="ECs1038" /note="similar to PABA_ECOLI gi 1787187 percent identity 100 in 172 aa (Conserved in E.coli K-12)" /codon_start=1 /evidence=not experimental /transl_table=11
CDS	gene	
	Query Match	99.6%; Score 454.4; DB 1; Length 327773;
	Best Local Similarity	99.8%; Pred No.7e-116; Mismatches 1; Indels 0; Gaps 0;
	Matches 455; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
Dy	1	ATCAAAAACAAATTGTTATTTTATGATTTTAACAATACTACTGGTCGCCCTGGATTGCAGCC 60
Dd	325925	ATGA AAAAACAATTGTTATTTATGATTTTAACAATACTACTGGTCGCCCTGGATTGCAGCC 325985
Dy	61	GCAGCAGGTTATGATTTAGCTAATTCAGAAATATACTCGCGGTAAATGAA TTGAGTAAG 120
Dd	325985	GCA GCAGGTTATGATTTAGCTAATTCAGAAATATACTCGCGGTAAATGAA TTGAGTAAG 326044
Dy	121	TCTTCATTATTCAGCGCACCATTAATGTCAAGCTGAGCACTAATAATAGTCTCAGTTA 180
Dd	326045	TCTTCATTATTCAGCGCACCATTAATGTCAAGCTGAGCACTAATAATAGTCTCAGTTA 326104
Dy	181	CGGACAGGAGGCTCAAAACCTTTTGGCGGTTGTGGCGCAAAAGTGTGACACCGGCGCA 240
Dd	326105	CGGACAGGAGGCTCAAAACCTTTTGGCGGTTGTGGCGCAAAAGTGTGACACCGGCGCA 326164
Dy	241	AAGATTGACCAGA CAGAGAGATTTAAACCTTGCACTATATTTGATGATGAGCGGCGAGTGCAC 300
Dd	326165	AAGATTGACCAGA CAGAGAGATTTAAACCTTGCACTATATTTGATGATGAGCGGCGAGTGCAC 326224
Dy	301	GATGCCAGTATTTGGCAAGTGCTTATGCTAATACTCGATGATTTATCAGAAAGTTCT 360
Dd	326225	GATGCCAGTATTTGGCAAGTGCTTATGCTAATACTCGATGATTTATCAGAAAGTTCT 326284
Dy	361	GGTAATAAAGCAAATTTATCACAGTATGCTCACTCAAAAAAGCGCAATTGTATGTGACAGA 420
Dd	326285	GGTAATAAAGCAAATTTATCACAGTATGCTCACTCAAAAAAGCGCAATTGTATGTGACAGA 326344
Dy	421	CAGTGCAGTAATGCTATTCGGGTGACACAAAGTTAA 456
Dd	326345	CAGTGCAGTAATGCTATTCGGGTGACACAAAGTTAA 326380
RESULT 7		
LOCUS	AE015131	10370 bp DNA linear BCT 18-OCT-2002
DEFINITION	Shigella flexneri 2a str. 301 section 94 of 412 of the complete genome.	
ACCESSION	AE015131 AE005674	
VERSION	AE015131.1 GI:24051313	
KEYWORDS		

SOURCE
ORGANISM
Shigella flexneri 2a str. 301
Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
1 (bases 1 to 10370)
Jin,Q., Yuan,Z.H., Xu,J.G., Wang,Y., Shen,Y., Lu,W.C., Wang,J.H.,
Liu,H., Zhang,X.B., Wu,H.T., Yang,F., Qu,D., Zhang,X.B., Zhang,J.Y., Yang,G.W.,
Mu,H.T., Dong J., Sun,L.L., Xue,Y., Zhao,A.L., Gao,Y.S., Zhu,Y.P.,
Kan,B., Chen,S.X., Yao,Z.J., He,B.K., Chen,R.S., Ma,D.L.,
Qiang,B.Q., Wen,Y.D., Hou,Y.D. and Yu,Y.D.
pathogenically through comparison with genomes of Escherichia coli
K12 and O157
Nucleic Acids Res. 30 (20), 4432-4441 (2002)

JOURNAL
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL

FEATURES
source
location/qualifiers
1..10370
/organism="Shigella flexneri 2a str. 301"
/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
complement(127..214)
/product="tRNA-Ser"
/note="anticodon: GGA"
410..1387
/gene="ycdW"
/locus_tag="SF1027"
/pseudo
410..1387
/gene="ycdX"
/locus_tag="SF1028"
/note="similar to Escherichia coli K12 putative
dehydrogenase gi: 1787270 (326 aa). BLAST with identity of
99% in 326 aa. This CDS contains an in-frame stop codon.
The sequence has been checked and is believed to be
correct."
/pseudo
/codon_start=1
/transl_table=1
1442..2179
/gene="ycdY"
/locus_tag="SF1028"
1442..2179
/gene="ycdX"
/locus_tag="SF1028"
/note="Residues 1 to 245 of 245 are 99 pct identical to
coli O157:H7 EDL933 ref: NP_287166.1"
/codon_start=1
/transl_table=1
/product="orf, conserved hypothetical protein"
/protein_id="AAK42652.1"
/db_xref="GI:24051314"
/translation="MYPVDLMHTVASTHAYSTLSPIYAOKGKIFALTDHGPD
MEAPHHMHIMRIRIPRVDSVGLRGIEINIKNGEIDCGKPFSDILDIATFHE
SLIHRKSSENCRCRAAARADAGCAALASDSTFTFTWGFEECLKTLDADVPLER
ILNVSRRLINFLRSRMAPIAEPAIDL"
2203..2757
/gene="ycdY"
/locus_tag="SF1029"
2203..2757
/gene="ycdY"
/gene="ycdY"

CDS
gene
CDS
gene
CDS
gene
CDS

```

/locus_tag="SF1029"
/note="Residues 1 to 184 of 184 are 99 pct identical to
residue 1 to 184 of a 184 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287169.1"
/codon_start=1
/transl_table=11
/product="putative oxidoreductase component"
/protein_id="AAN42653.1"
/db_xref="GI:24051315"
/translaction="MNEPSIILCRVLGSLVYRQDPDLVLPFTLIRRGKLANWPLEQ
DELLRLQGSCKMTQVSADYVALFGDECAVPPYRASVEDATAEAVRAPISEGGML
ADTPDHIGTLLASMTQSTEDSESALETLSEVILLPMCGANFLKVEAHATTPTPW
RTMAULPTDAISAMDELEBDESE"
2811..3350
/gene="ycd2"
/locus_tag="SF1030"
2811..3350
/gene="ycd2"
/locus_tag="SF1030"
/note="Residues 1 to 179 of 179 are 98 pct identical to
residue 1 to 179 of a 179 aa protein from Escherichia
coli K12 ref: NP_415554.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AAN42654.1"
/db_xref="GI:24051316"
/translaction="MCSISPMAPASAIIRGNMILLIAITTGILSGIMGWAVSLGL
SWAGLGTGAVPACPOGGLKGLIAALTLGSGVWVWVVIYGSALAPHEILICVITG
IVAFMCIQAOAILLSFVPGTFIGACATFAGCGMKVLVPSIALGLVFGYAMKNSGIM
LAARSAKTAHNEQETKKA"
complement(3414..4189)
/gene="csgC"
/locus_tag="SF1031"
/pseudo
complement(3414..4189)
/gene="csgC"
/locus_tag="SF1031"
/note="similar to Escherichia coli K12 curI production
assembly-transport component, 2nd curI operon gi:1787274
(1278 aa). BLAST with identity of 97% in 257 aa. This CDS
contains an in-frame stop codon. The sequence has been
checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=11
complement(4454..5662)
/locus_tag="SF1032"
complement(4454..5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to
residue 1 to 402 of a 402 aa protein from Plaemid R100
ref: NP_052925.1"
/codon_start=1
/transl_table=11
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translaction="WCELDIILHDSLYOPCEPLHLKRLNSLTACLHALDOCKTITLTEL
GNSLPTKARKTANIKRIDRLGNHLEHKEKRLAVIRHASFCSGNTWPIYLVWMSDIR
EOKRLMVIARSAVALHGRSVTLYEKAFPLSEQCSKKAHDOFLADLASILPENTPLIYS
DAGFVPVYKSGVEKGLKWSRVRGKVQYADLGAENMKPISNLHDMSSSHSKTLGYKR
LTKSPILSCQILLYKSRGKRNORSTRHCHBSPTIYASAKPEITATNLPVEIR
TQKLVNIYSRKMOIEETFFDLKSPAYGLGRHSRTSSSRPDMILLIAMLQTCML
AGVAKORGNKHKHQANTVVRNRYNLSTVRLGMEVLRHSQTTITREDSVAATITLTONL
FTHGVLVLEKL"
complement(5995..6384)
/gene="csgE"
/locus_tag="SF1033"
complement(5995..6384)
/gene="csgE"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to

```

residues 1 to 129 of a 129 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP_287173.1"

1
/codon_start=1
/transl_table=11
/product="curli production assembly-transport component,
2nd curli operon"
/protein_id="AA042656.1"
/db_xref="GI:24051318"
/translation="MKRYRWIYAAEFLPAAGNHAWEVEVPGILTDHTVSSIGHDFY
RAFDKESDYTGNIILINERSPARSGSWITITVNQVIFQTFPLPKRDEKTVFAL
IQRFEALNRQINQALISTDDLAHDE"

gene
complement(6389. 7069)
/gene="csgd"
/locus_tag="SF1034"
/locus_tag="SF1034"
complement(6389. 7069)
/gene="csgd"
/locus_tag="SF1034"
/notes="Residues 11 to 226 of 226 are 100 pct identical to
residues 1 to 216 of a 216 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=11
/product="putative 2-component transcriptional regulator
for 2nd curli operon"
/protein_id="AA042657.1"
/db_xref="GI:24051319"
/translation="MRSIKKGFIMFENEVHSIGHTLTLITPKSLQATALLKQSL
AIFGKHLNIDRSIDLSGSSITLLDMMADAKLIHWODTLSRKNNIKILINPTED
YPRDLENPHIQRSDVFPAMBDORVYNGLOSVIRGECYPTQKSLAIVTHSGRYNS
TESALLTHREKEILINKLRIGASNNELARSFISENTVTKLIVLFFKILAKNRTQAVS
WADNDLNR"

gene
7766. 8248
/gene="csgb"
/locus_tag="SF1035"
7766. 8248
/gene="csgb"
/locus_tag="SF1035"
/notes="Residues 10 to 160 of 160 are 99 pct identical to
residues 1 to 151 of a 151 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP_287175.1"
/codon_start=1
/transl_table=11
/product="minor curli subunit precursor, similar to CsgA"
/protein_id="AA042658.1"
/db_xref="GI:24051320"
/translation="MYDOYGDGMKKKLLFPMULTIIGAPGIAAAGYDLANSRYNAV
NEISKSPNOALITIGAGTNSKQLROGGSKLLAAYAOBSSSRPAKIDQGDYNIAYI
DQGSANDASISQAGYNTAMIIQKSGSNKANTYGTQTAIVVQRQSOAMLRVTR

Query Match 98.9%; Score 451.2; DB 1; Length 10370;
Best Local Similarity 99.3%; Pred. No. 5,8e-115;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 ATGAAAAACAATGTTATTATGATGTTAAACAATCTGGGTCGCTGGGATTCAGCC 60
7793 ATGAAAAACAATGTTATTATGATGTTAAACAATCTGGGTCGCTGGGATTCAGCC 7852
61 GCGCAGGTTATGATTTAGCTAATTGAGATATTAACCTTCGCGGTTAATGAAATTGAGTAAG 120
7853 GCGCAGGTTATGATTTAGCTAATTGAGATATTAACCTTCGCGGTTAATGAAATTGAGTAAG 7912
121 TCTTCATTATATCAGGAGCATATATGTCACGCTGGGACTAATATAGTCTCAGTTA 180
7913 TCTTCATTATATCAGGAGCATATATGTCACGCTGGGACTAATATAGTCTCAGTTA 7972
181 CCGCAGGAGGCTCAAAAATCTTTGGCGGTTGTCGCCAAGAGGTAGTACAAACCGGCA 240
Db CCGCAGGAGGCTCAAAAATCTTTGGCGGTTGTCGCCAAGAGGTAGTACAAACCGGCA 8032
7973 CCGCAGGAGGCTCAAAAATCTTTGGCGGTTGTCGCCAAGAGGTAGTACAAACCGGCA 8032
241 AAGATTGACGACGACGAGATTTAATCCTTGCAATATATGATACGCGGCAATGGCAAC 300
Db AAGATTGACGACGACGAGATTTAATCCTTGCAATATATGATACGCGGCAATGGCAAC 8092


```
/product="Putative conserved protein"
/protein_id="AAN79731.1"
/db_xref="GI:26107533"
/translation="MPROPDGYREIKOTMOLSLTAKSKPELVGEELKAVLLAAYNL
VRYQMIKAEHLKGVNQLSPFSCGVMMLMTLQASPGRIPELMRDILASMGOLV
KLPTTRREAPPRVVERP"
7651..8013
/locus_tag="c1258"
7651..8013
/locus_tag="c1258"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AAN79732.1"
/db_xref="GI:26107533"
/translation="MLGACFTSMGTSFTYGFPPMGAELVLMMPBACMLLIQSLYSL
LDGEQVLFSEWLNKKQYEGFEFYQIFSGNIVASCSCSVSSNSAHYACDFTMRH
YSLNNYACQLGRACLS"
7835..8269
/locus_tag="c1259"
7835..8269
/locus_tag="c1259"
/locus_tag="c1259"
/note="Residues 1 to 139 of 144 are 50.71 pct identical to
residues 104 to 242 of 258 from Genpept.129 :
>gb|AAK18487.1|AF348706_176 (AF348706)orf, hypothetical
[Shigella flexneri]"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AAN79733.1"
/db_xref="GI:26107534"
/translation="MKDLSPFRPLALLLPALPEPVIRHRTGALTLCTYTL
ITTLASGRLVSKCISESPTIRRHIECMAGFIYFTFSLIAGWYRQKKEY
OTLYENRKYVLYARDGLVLSQSYRSGRKRVYINGHRL"
8477..8553
/locus_tag="c5534"
8477..8553
/locus_tag="c5534"
/product="rRNA-Arg"
/function="rRNA; rRNA"
/complement(8832..9290)
/locus_tag="c1260"
/complement(8832..9290)
/locus_tag="c1260"
/function="putative enzyme; Transposases Insertion
Sequence Associated"
/note="Escherichia coli O157:H7 ortholog: z4324"
/codon_start=1
/transl_table=1
/product="Putative Transposase"
/protein_id="AAN79734.1"
/db_xref="GI:26107535"
/translation="MASAADARIVSHYERMLIEYHKAWKSGTCVESLRMOTRDN
LERNVVIAPIAVMLGLRQGISSETQNDCKILPTTEWKLIMVKLEGQLPSQTP
TLKAACTLGRHDSKRTGRPMVVMMDGFRLODMVGYVMSLDEI"
complement(9302..9667)
/locus_tag="c1261"
complement(9302..9667)
/locus_tag="c1261"
/complement(9302..9667)
/note="Residues 1 to 35 of 121 are 97.14 pct identical to
residues 59 to 93 of 316 from Genpept.129 :
>gb|AAK18437.1|AF348706_126 (AF348706) putative
transposase [Shigella flexneri]"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AAN79735.1"
/db_xref="GI:26107536"
/translation="MAERLGRIOKRVITVCREADIMVLYLHYKXSHGCRFVVRTAQNS
```

```
Query Match 98.9%; Score 451.2; DB 1; Length 306358;
Best Local Similarity 99.3%; Pred. No. 5.5e-115;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 1 ATGAAAAACAATGTTATTTATGATGTTAACAATTAAGTGGGCGCTGGGATTCACACC 60
Db 44773 ATGAAAAACAATGTTATTTATGATGTTAACAATTAAGTGGGCGCTGGGATTCACACC 44832
QY 61 GCACAGAGTTATGATTTAGTAAATTCAGATATTAATTCGGGGTAAATGAATAGTAAAG 120
Db 44833 GCACAGAGTTATGATTTAGTAAATTCAGATATTAATTCGGGGTAAATGAATAGTAAAG 44892
QY 121 TCTTCATTTAATCAAGCAGCCATTAATGTCAGCTGGGACTAATATAGTGCTCAGTTA 180
Db 44893 TCTTCATTTAATCAAGCAGCCATTAATGTCAGCTGGGACTAATATAGTGCTCAGTTA 44952
QY 181 CGCAGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTATGACACCGGGCA 240
Db 44953 CGCAGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTATGACACCGGGCA 45012
QY 241 AAGATTGACAGACAGGAGATTATTAACCTTGATATTTATGATCAGCGGCGAGTCCAAC 300
Db 45013 AAGATTGATCAGACAGGAGATTATTAACCTTGATATTTATGATCAGCGGCGAGTCCAAC 45072
QY 301 GATGCCAGTATTTGCGAAGGTGCTTAATGATATTAATCTGGATGATTAATCAGAAAGTTCT 360
Db 45073 GATGCCAGTATTTGCGAAGGTGCTTAATGATATTAATCTGGATGATTAATCAGAAAGTTCT 45132
QY 361 GGTATTAAGCAATATTTACAGTATGATCTCAAAAAGCGCAATTTGATGACAGGA 420
Db 45133 GGTATTAAGCAATATTTACAGTATGATCTCAAAAAGCGCAATTTGATGACAGGA 45192
QY 421 CAGTCGCAATGCGCTATTCGCGTACACCAACGTTAA 456
Db 45193 CAGTCGCAATGCGCTATTCGCGTACACCAACGTTAA 45228

RESULT 9
AE016981 292504 bp DNA linear BCT 22-APR-2003
LOCUS Shigella flexneri 2a str. 2457T section 4 of 16 of the complete
DEFINITION genome.
ACCESSION AE016981 AE014073
VERSION AE016981.1 GI:30040616
KEYWORDS
SOURCE
ORANISM
Shigella flexneri 2a str. 2457T
Shigella flexneri 2a str. 2457T
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE
1 (bases 1 to 292504)
Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,M.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Complete Genome Sequence and Comparative Genomics of Shigella
flexneri Serotype 2a Strain 2457T
Infect. Immun. 71 (5), 2775-2786 (2003)
12704152
2 (bases 1 to 292504)
Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,M.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Direct Submission
Submitted (13-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1..292504
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
/complement(223..441)
/gene="infA"
/locus_tag="S0884"
```



```

CDS
    complement(223..441)
    /gene="infA"
    /locus_tag="S0884"
    /function="factor; Proteins - translation and modification"
    /note="residues 1 to 72 of 72 are 100.00 pct identical to residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
    /codon_start=1
    /transl_table=11
    /product="protein chain initiation factor IF-1"
    /protein_id="AAP16348.1"
    /db_xref="GI:30040617"
    /translation="MAKEDNIMOGTYLITLPTMFWELNENHVVTAHISGRKRY
    IRLTGDKYVVELTPYDLSKRIYFRSR"
    complement(726..1430)
    /gene="at"
    /locus_tag="S0885"
    complement(726..1430)
    /gene="at"
    /locus_tag="S0885"
    /function="enzyme; Aminoacyl tRNA synthetases, tRNA modification"
    /note="residues 1 to 234 of 234 are 100.00 pct identical to residues 1 to 234 of 234 from Escherichia coli K-12 : B0885"
    /codon_start=1
    /transl_table=11
    /product="leucyl, phenylalanyl-tRNA-protein transferase"
    /protein_id="AAP16349.1"
    /db_xref="GI:30040618"
    /translation="KRLVQLSRHSIAFPSPGALREPNGLALGDLSPARLLMAYOR
    GIPFSPGDPILMSPDPRAVLMPESLHISMSKRFKRSRYVTMYAAGVITBC
    ASDREGVITRGVAVAHRLHGLHSAHSEVREDELVGWVAGVAGCTLECGSMS
    RMENASKRLALVCEEFIGHGKGLDCVLDNDHTASLACIPPRDYIANTYINQRLSR
    LNNFWVPKCLSPDE"
    complement(1472..3193)
    /gene="cydC"
    /locus_tag="S0886"
    complement(1472..3193)
    /gene="cydC"
    /locus_tag="S0886"
    /function="Transport of small molecules: Other"
    /note="residues 1 to 573 of 573 are 97.03 pct identical to residues 1 to 573 of 573 from Escherichia coli K-12 : B0886"
    /codon_start=1
    /transl_table=11
    /product="ATP-binding component of cytochrome-related transport"
    /protein_id="AAP16350.1"
    /db_xref="GI:30040619"
    /translation="MRALLPYALLYKHKMLSLGIVLAIVTLASIGLLTSGWFLS
    ASAVAGVGLSYFMYLPAAGVGAATTRNGVFERLVSHDAPFRVLQHLTYFSK
    LPLSPAGLARYGSELINRVADVDLDLHYRIVSPVGAFFVIMVTLISPLFD
    LTATIGSLWTLITFLMPPLFRAGSSTGNTLTHRGVQQLTAMLGQAALITFGA
    SDARTLENTETIOWLEAQRROSELTALSCAIIWLIGALAVITLMMSCGGAGACP
    GALRALVFCALAAFEALAVGAFQHLQVIVASAVRISDLTDOKPEVTPPTQVRA
    DRSVLTLDVQVTPYEPESQALKGISLOVAGAHIAIIGRTGCGSTLQTLTRAWD
    QOGEILNDSPILASINEALRORTISVAPORVHLFSTILFDNLILASPESSDEALSEIL
    RRVGLEKLEEDAGNSWIGEGGROISGSELRLATRLHDAPLVLLDEPTEGIDAT
    TESQIETLAKMKREKTVLMTYHRLRGLSRFOIITVMDNGOIIEOGTAEHLAAGRY
    YORKQGI"
    complement(3194..4960)
    /gene="cydD"
    /locus_tag="S0887"
    complement(3194..4960)
    /gene="cydD"
    /locus_tag="S0887"
    /function="transport; Not classified"
    /note="residues 1 to 588 of 588 are 92.85 pct identical to residues 1 to 588 of 588 from Escherichia coli K-12 : B0887"
    /codon_start=1

gene
    /transl_table=11
    /product="ATP-binding component of cytochrome-related transport"
    /protein_id="AAP16351.1"
    /db_xref="GI:30040620"
    /translation="MNRKROKELPRMLKOOSVTSQWMLNIRLIGVSGIILTAQAMF
    MARLIQHMENIIPREALLPPTLVITPTFLRAVYVWLEBRYGHAQHIRPAIRQV
    LDRLOAGPAMIOGKPPAGSWATVLEQIDMDHYARYLPQMAVAVPLLIIVAFIP
    SMMVAALILITGAPDILPRLVVGGAADNRNPLALRSGHFLDRLRGMEYIRIF
    GGRGARETSRSASSEDPRQRTMEVRLAEISGIIFFTSLSIALVAVFGSYGEL
    DRGYDTGVLTAAGFLALIAPEFQPLRDLGTFYAKAQAAGADSLKTEMEPLAH
    PQSEBALUTDPLITTEADLFTSPGKTLAGPLNFTLPAGORAVIVAVSSGSSSL
    LNALSGLSYQSLIRINGIELRDSERSKHLISWYGNPQIPAKKDNYVALPDA
    SEQELQALDNWAVSEFLPLDPGVDTPIGDOAAULSVCORVAVAPALNPSLL
    LDEPAAISIDHSEORVWELNALNASLRQTLMTYHQLELDMDVITWQDRRIIEGR
    YAEISVAGGPFATILAHROET"
    complement(5083..6048)
    /gene="trxB"
    /locus_tag="S0888"
    complement(5083..6048)
    /gene="trxB"
    /locus_tag="S0888"
    /function="enzyme; 2-Deoxyribonucleotide metabolism"
    /note="residues 1 to 321 of 321 are 95.95 pct identical to residues 1 to 321 of 321 from Escherichia coli K-12 : B0888"
    /codon_start=1
    /transl_table=11
    /product="chlorodoxin reductase"
    /protein_id="AAP16352.1"
    /db_xref="GI:30040621"
    /translation="MGTTKISKILLDSGPRAGTAAYARANILPVLITGMEKGGQL
    TTTEVENMPDPPDLTGPILMARHSHATKFTETLIPDIINKVDQNRPFRLNGNG
    EYTCDDALITAGASARVIGLSEAEFGKQVSACTDGFYNOQVAVYAGGNTAVE
    EALVTSINASVHLIHRDPRAEKILIKIMDKVKNNGNIIHTNTLEEVGDMQGV
    TGVRLQDONSNDIESLDVAGLFAVIGHSPTAIIPGQLELENGYIKYKSHGNATQ
    TSIPGVPAAGDVMDHIYRQAITSGTCMAALDAENYLDGLADAK"
    6592..7086
    /gene="lrp"
    /locus_tag="S0889"
    6592..7086
    /gene="lrp"
    /locus_tag="S0889"
    /function="regulator; Central intermediary metabolism: Pool, multipurpose conversions"
    /note="residues 1 to 164 of 164 are 100.00 pct identical to residues 1 to 164 of 164 from Escherichia coli K-12 : B0889"
    /codon_start=1
    /transl_table=11
    /product="regulator for leucine (or lrp) regulon and high-affinity branched-chain amino acid transport system"
    /protein_id="AAP16353.1"
    /db_xref="GI:30040622"
    /translation="MWSKRRPKRDLRIDRNNITNEIQKGRISNVELSRVGLSPRP
    CIERVRLRBOGFOGTALNRPYLDLSLIVETILNRGADVPFOFTVAQKLE
    IOEHLVSGDDPYLTKTVPMASARKLIGETLLRLPGVNDTRYVMEVKOSNRLY
    ITRR"
    7221..11249
    /gene="ftsK"
    /locus_tag="S0890"
    7221..11249
    /gene="ftsK"
    /locus_tag="S0890"
    /function="phenotype; Cell division"
    /note="residues 1 to 737 of 1342 are 68.24 pct identical to residues 1 to 737 of 1329 from Escherichia coli K-12 : B0890"
    /codon_start=1
    /transl_table=11
    /product="cell division protein"
    /protein_id="AAP16354.1"
    /db_xref="GI:30040623"

```


/translation="MSOEYEDKEVTLTKLSSGRLLBALLLIVLFAVWMAALLSF
 NSPDSFSGTAMHEPIHNLGGWPMGLADLTLEFIFGWAAATTPVIVGGCFNAPHOS
 SDEYIVFAVSLRITIGVALITLSCGLAJNADITWYASATTPVIGSGNAPRHHOS
 SGGTIALLCWAAGLTLETTGMSWTIARKLGMILNITPFSNRKRDITWDEDEYE
 DDEEYEDENHKGHESRRARILRGLARKRLAEKFINPMKQTDALFSGRMDDE
 EITTYARGVADPDVPLSGNRAITQPEYEDVPLNSVPIEPVAVAAATTAOTSM
 APEVETQTPVAVSPQPTVAMQVPGQTEPIYADAPEGYPOOSQACQAVQ
 YNEPLQOPVQOQPYAPAAEQPAQPYAPAEQPVAGNMAOEQOSTPASTVO
 TEOTYQGPAAOEPLVQOPVQOPVPEPVEPTKAPPLIYFEEVEERARERE
 OLAAVQPIPEPVEPPEIKSLKPSAAVPPVAAAASVPLASGVKATLATGAA
 TVAAVFSLSNGSRPQVKEGIGQLPRPRIIVPTRELASVGIKLPSQRADEKA
 REAQRNOYSDGVNDDIDAMQODELROFQOQVGEYOHVNVNEDADA
 EAEIARQPAQTOOQVSGEOPAGANFSLDFEFSMKALLDDGHEPLFTPIVQ
 QOQPVAPQOQVQOQVAPQOQVQOQVAPQOQVAPQOQVAPQOQVAPQOQV
 APQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQV
 DLTPRESEVPEVDFALBQMARLEALBAPRIADVNTSPCVITREFNLAPGV
 KAARISNLRDLARSLTVAARVAVVIEGKPVGELPKKRQTVYLAEVLDNAKFRD

Query Match 98.6%; Score 449.6; DB 1; Length 292504;
 Best Local Similarity 99.1%; Pred. No. 1.5e-114;
 Matches 452; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAATTTTATGATGTTAACAATATCTGGGTCGCGCTGGATTGCACCC 60
 Db 210815 ATGAAAAACAATTTTATGATGTTAACAATATCTGGGTCGCGCTGGATTGCACCC 210874
 Qy 61 GCAGAGGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGATAG 120
 Db 210875 GCAGAGGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGATAG 210934
 Qy 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAA 180
 Db 210935 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAA 210994
 Qy 181 CGGCAAGGAGGCTCAAAACCTTTGGCGGTGTGGCGCAAGAGGTAGTACGACCGGGCA 240
 Db 210995 CGGCAAGGAGGCTCAAAACCTTTGGCGGTGTGGCGCAAGAGGTAGTACGACCGGGCA 211054
 Qy 241 AAGATTGACGAGCAGGAGATTAATTAATCTTCGATATTTGATCAGCGCGGCGACGTGCCAAC 300
 Db 211055 AAGATTGACGAGCAGGAGATTAATTAATCTTCGATATTTGATCAGCGCGGCGACGTGCCAAC 211114
 Qy 301 GATGCCAGTATTTGCAAGTGTCTTATGTAATCTGCGATGATTTCCAGAAAGGTTCT 360
 Db 211115 GATGCCAGTATTTGCAAGTGTCTTATGTAATCTGCGATGATTTCCAGAAAGGTTCT 211174
 Qy 361 GGTATATAAGCAAAATATTACACAGTATGTAATCTCAAAAACCGCAATTGTAGTGCAGAGA 420
 Db 211175 GGTATATAAGCAAAATATTACACAGTATGTAATCTCAAAAACCGCAATTGTAGTGCAGAGA 211234
 Qy 421 CAGTCGCAATGCGTATTCGGGTGACACAACTTAA 456
 Db 211235 CAGTCGCAATGCGTATTCGGGTGACACAACTTAA 211270

RESULT 10
 CSEP515700 2889 bp DNA linear BCT 24-JUN-2003
 LOCUS CSEP515700 2889 bp DNA linear BCT 24-JUN-2003
 DEFINITION Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
 ACCESSION AJ515700
 VERSION AJ515700.1 GI:31790491
 KEYWORDS csgB gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
 component of curlin monomers; regulatory protein.
 SOURCE Citrobacter sp. Fec2
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Citrobacter.
 REFERENCE 1
 AUTHORS Zogaj X., Bokranz W., Nimtz M. and Bowling U.
 TITLE Production of Cellulose and Curli Fimbriae by Members of the Family
 JOURNAL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
 REFERENCE 2 (bases 1 to 2889)

AUTHORS Romling U.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
 Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES location/Qualifiers
 source 1..2889

/organism="Citrobacter sp. Fec2"
 /mol_type="genomic DNA"
 /isolate="Fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /country="Germany"
 /complement(121..862)
 /gene="csgB"
 /complement(212..862)
 /gene="csgD"
 /function="necessary for transcription of the csgA
 operon"

/codon_start=1
 /transl_table=1
 /product="putative regulatory protein"
 /protein_id="CAD56670.1"
 /db_xref="GI:31790492"

/translation="MFNEPFSIHGHTLLITKPSLOAPALLOHLKOSLALSGKLNTIO
 RSLDDISSCIVVDMMEDAKLIHYMDNISRNKLKTLTTPDTPYRDLNMP
 HINQVYADDERVNVNGLOITLGEFCFSKLSYLITHSNTRINSSSALLTHRS
 KEILNKRIGASNEIARSLFISNTVTKHLYNFKIYAVNRTOAVSWANDLRR"
 /length=2078
 /gene="csgB"
 /length=2078
 /gene="csgB"
 /function="nucleator"

/codon_start=1
 /transl_table=1
 /product="nucleation component of curlin monomers"
 /protein_id="CAD56671.1"
 /db_xref="GI:31790493"

/translation="MKRLQVAFALIVSGSALAGVDPWQGGGGGGSSGSESTL
 QALITGVGNNSAKMDEGSKLSVVSQEGSRARVDSGAVNFAVIAQSGHNSDA
 SISQNSGNTAMIIQKSGNKANITQYGTOKTAVVOROSMAIRVIO"
 /length=2571
 /gene="csgA"
 /length=2571
 /gene="csgA"
 /function="curli subunit"

/codon_start=1
 /transl_table=1
 /product="curlin-csgA protein"
 /protein_id="CAD56672.1"

/db_xref="GI:31790494"
 /translation="MKLLQVAFALIVSGSALAGVDPWQGGGGGGSSGSESTL
 SIQVGNNAALALQSDARKSDITIHQNGGNGADVQGSNSSTIDITONGFNKNATI
 DQWGNKSDITVQYQGHNAALVNGTASDSVLVHQVGFNNATANY"

ORIGIN

Query Match 78.2%; Score 356.8; DB 1; Length 2889;
 Best Local Similarity 86.4%; Pred. No. 1.3e-88;
 Matches 394; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
 Qy 1 ATGAAAAACAATTTTATGATGTTAACAATATCTGGGTCGCGCTGGATTGCACCC 60
 Db 1623 ATGAAAAACAATTTTATGATGTTAACAATATCTGGGTCGCGCTGGATTGCCTTC 1682
 Qy 61 GCAGAGGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGATAG 120
 Db 1683 GCAGAGGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGATAG 1742
 Qy 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAA 180
 Db 1743 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTAA 1802
 Qy 181 CGGCAAGGAGGCTCAAAACCTTTGGCGGTGTGGCGCAAGAGGTAGTACGACCGGGCA 240


```

Db      996  GATGCACATATATGCGCAAGCGCTTACGCTATAGTACGACGCTATATCCAGAAAAGTCT 1055
Oy      361  GGTATTAAGCAATATTTACACAGTATGTTACTCAAAAACGGCAATTGTAGTCAGAGA 420
Db      1056  GGAATATAGGCGCAATATTACCAGTACCGTACGCGAAGAACGACAGTGTAGTCAGAAA 1115
Oy      421  CAGTCGCAATGGCGCTATTGGGTGACACAGCTTAA 456
Db      1116  CAGTCGCAATGGCTATTGGCTCCACCAAGCTTAA 1151

RESULT 12
STAD301 5103 bp DNA linear BCT 15-NOV-2000
LOCUS Salmoneilla typhimurium cs9g, cs9f, cs9g, cs9d, cs9b, cs9a, and k
DEFINITION
ACCESSION AJ002301
VERSION AJ002301.1 GI:2739232
KEYWORDS cs9a gene; cs9b gene; cs9c gene; cs9d gene; cs9e gene; cs9f gene;
SOURCE Salmoneilla typhimurium
ORGANISM Salmoneilla typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmoneilla.
REFERENCE 1
AUTHORS Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Normark, S.
TITLE Culi fibers are highly conserved between Salmoneilla typhimurium
and Escherichia coli with respect to operon structure and
regulation
JOURNAL J. Bacteriol. 180 (3), 722-731 (1998)
MEDLINE 9457880
PUBMED 9457880
REFERENCE 2 (bases 1 to 5103)
AUTHORS Romling, U.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES
source
1..5103
Location/Qualifiers
/organism="Salmoneilla typhimurium"
/mol_type="genomic DNA"
/strain="SR-11"
/db_xref="taxon:602"
<1..322
/note="ORF179a"
/codon_start=2
/transl_table=1
/product="hypothetical protein"
/protein_id="CA05311.1"
/db_xref="GI:2739233"
/db_xref="GOA:O54290"
/db_xref="SWISS-PROT:O54290"
/translation="CTLSGWMALVLIHGSALAPHLIEIVSYLTGVIAPLMCIQAKO
LLSFVPGTFIGACATFAGCGDMRLVLPGLGLIFGVAMKNSGLMLASRRREGSANT
AVTK"
complement(361..1194)
/gene="cs9g"
complement(361..1194)
/gene="cs9g"
/function="lipoprotein"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Cs9g protein"
/protein_id="CA05312.1"
/db_xref="GI:2739234"
/db_xref="GOA:O54291"
/db_xref="SWISS-PROT:O54291"
/translation="MPRLILVAALVLSGCCITAPPKAKFTLMPRAQSYKDLTHLPA
PTGKIFVSVNIQDETQCFKPYPSNSFTAPQATMLVTALKDSRFILPLEQGLQ
NLNBRKIRAAQENGTAVMNNRIPLQSLTFAANIIVSGSLIGESNVKSGGVARYG
IGADTOYQLDQIAVNLRVNVSTGEILSVTSKTIISYEVQAGVPRFIDYQRLBEE

```

```

gene
complement(1221..1637)
/gene="cs9f"
complement(1221..1637)
/gene="cs9f"
CDS
/codon_start=1
/transl_table=1
/product="Cs9f protein"
/protein_id="CA05313.1"
/db_xref="GI:2739235"
/db_xref="SWISS-PROT:O54292"
/translation="MRVKAHVLIIMLFSPLTWAGMTPFPNPNRGGNPNNGSPLNS
AQONSXKDPAYNDPBGIERPSALDNFTQAIQSILGGLTINITGARGAVTWDFTI
DIARDQQLQUNVDRKTRSTIEVGLQSTDP"
complement(1664..2059)
/gene="cs9b"
complement(1664..2059)
/gene="cs9b"
CDS
/codon_start=1
/transl_table=1
/product="Cs9b protein"
/protein_id="CA05314.1"
/db_xref="GI:2739236"
/db_xref="SWISS-PROT:O54293"
/translation="MKRYLTWYVAELLPATGNLHANEVEVPEGLTDHTVSIGHE
FYRAFSDKWSSEYTGNTINERPARGWSITTVNODVIFQTLFPKQDFEKTIVF
ALAQTEALRLRRQIDQTLSTSLARDF"
complement(2064..2714)
/gene="cs9d"
complement(2064..2714)
/gene="cs9d"
CDS
/function="transcriptional regulator"
/codon_start=1
/transl_table=1
/product="Cs9d protein"
/protein_id="CA05315.1"
/db_xref="GI:2739237"
/db_xref="GOA:O54294"
/db_xref="SWISS-PROT:O54294"
/translation="MFNEVHSHGHITLLITKPSIQATLALQHLKQSLATNGKLANIQ
RSLEIDBAGCIVLMDMEADKLIHYQDNLSRKNNIKITLLNTPDDYPYRIENNP
HINGVFATGSDQEHVSGLOGIGECYFQOKLASYLITSGNVRVSTSLATLTHE
KEILNKIRICASNNETARSFISNTYKTHLYLFKLIVAKRTQAVNSMNDINLR"
complement(2869..2896)
/gene="cs9DERG"
complement(2869..2874)
/gene="cs9DERG"
complement(2891..2896)
/gene="cs9DERG"
3344..3376
/gene="cs9BA"
3344..3349
/gene="cs9BA"
3371..3376
/gene="cs9BA"
3469..3924
/gene="cs9B"
3469..3924
/gene="cs9B"
/function="nucleator"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Cs9b protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translation="MKRKLFLPMULTIIGAGCITATVATVILARBEYNFANVLSKSN
QAAIIGVGTGDNARVROEGSKLSVSGCGNNRAVDAGNVAIFVIEOTGAND
SISQASGNSAALIIQKSGKKNKITVGTGKTAVVVGKQSHMAIRVQR"
3469..3531
sig_peptide

```

```

mat_peptide      /gene="csaB"
                  3532..3921
                  /gene="csaB"
                  /product="CsaB protein"
                  /function="nucleator"
gene             3966..4421
                  /gene="csaB"
                  /gene="csaB"
                  /function="fiber subunit"
                  /codon_start=1
                  /evidence-experimental
                  /transl_table=11
                  /product="CsaB protein"
                  /protein_id="CA05317.1"
                  /db_xref="GI:2739239"
                  /db_xref="GOA:P55225"
                  /db_xref="SWISS-PROT:P55225"
                  /translation="MKLLKVAFAIVAGSALAGVVPQMGNGGNGGNSGPDST
                  IDQMAKNSDITVGQYGGNNAALVNGTASDSVWVRQVGFNNATANY"
CDS              3966..4025
                  /gene="csaB"
                  /gene="csaB"
                  /product="CsaB protein"
                  /function="fiber subunit"
                  /transl_table=11
                  /codon_start=1
                  /product="CsaB protein"
                  /protein_id="CA05318.1"
                  /db_xref="GI:2739240"
                  /db_xref="SWISS-PROT:P55227"
                  /translation="MTLLLLAALNSQTFPTTQGGDIYTVIPQVTLNEPCQVOYL
                  WPSAQ"
sig_peptide      4483..4533
                  /gene="csaB"
                  /gene="csaB"
                  /product="CsaB protein"
                  /function="fiber subunit"
                  /transl_table=11
                  /codon_start=1
                  /product="CsaB protein"
                  /protein_id="CA05319.1"
                  /db_xref="GI:2739241"
                  /db_xref="SWISS-PROT:P55228"
                  /translation="MSVYIKKNIPALIGLCAPFIHSAVGQGVVHFGAIVEPL
                  CDISTHAENI"
gene             4534..4806
                  /gene="csaB"
                  /product="CsaB protein"
                  /function="fiber subunit"
                  /transl_table=11
                  /codon_start=1
                  /product="CsaB protein"
                  /protein_id="CA05320.1"
                  /db_xref="GI:2739242"
                  /db_xref="SWISS-PROT:P55229"
                  /translation="MSVYIKKNIPALIGLCAPFIHSAVGQGVVHFGAIVEPL
                  CDISTHAENI"
CDS              4807..5103
                  /gene="csaB"
                  /product="CsaB protein"
                  /function="fiber subunit"
                  /transl_table=11
                  /codon_start=1
                  /product="CsaB protein"
                  /protein_id="CA05321.1"
                  /db_xref="GI:2739243"
                  /db_xref="SWISS-PROT:P55230"
                  /translation="MSVYIKKNIPALIGLCAPFIHSAVGQGVVHFGAIVEPL
                  CDISTHAENI"

```

ORIGIN

Query Match 73.0%; Score 332.8; DB 1; Length 5103;
 Best Local Similarity 83.1%; Pred. No. 6.3e-82;
 Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```

QY 1 ATGAAAACAAATTTTATTGATGTTAACAATACGCGGCGCCGCGATTCGACCC 60
DB 3469 ATGAAAACAAATTTTATTGATGTTAACAATACGCGGCGCCGCGATTCGACCC 60
QY 61 GCAGCAGGTATGATTTAGTCAATTAATCTTCGCGTAAATGAATGATGATG 120
DB 3529 GCAGCAGGTATGATTTAGTCAATTAATCTTCGCGTAAATGAATGATGATG 120
QY 121 TCTTCAATTAATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3588
DB 3589 TCTTCAATTAATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 3648

```

```

QY 181 CGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 240
DB 3649 CGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3708
QY 241 AAGATTGACACGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3708
DB 3709 AAGATTGACACGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 3768
QY 301 GATGCGATTTTGGCAGAGTCTTATGATGATGATGATGATGATGATGATGAT 360
DB 3769 GATGCGATTTTGGCAGAGTCTTATGATGATGATGATGATGATGATGATGAT 3828
QY 3829 GGAATTAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420
DB 421 GATGCGATTTTGGCAGAGTCTTATGATGATGATGATGATGATGATGATGAT 3888
QY 3889 GATGCGATTTTGGCAGAGTCTTATGATGATGATGATGATGATGATGATGAT 3924

```

RESULT 13
 AE008749 22411 bp DNA linear BCT 23-Apr-2003
 LOCUS Salmonella typhimurium LT2, section 53 of 220 of the complete
 ACCESSION AE008749 AE006468
 VERSION AE008749.1 GI:16419641
 KEYWORDS
 SOURCE
 ORGANISM
 Salmonella typhimurium LT2
 Salmonella typhimurium LT2
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.
 1 (bases 1 to 22411)
 McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*
 LT2
 Nature 413 (6858), 852-856 (2001)
 MEDLINE 21534948
 PUBMED 11677609
 2 (bases 1 to 22411)

REFERENCE
 AUTHORS
 CONSRM
 TITLE
 JOURNAL
 COMMENT
 The Salmonella typhimurium Genome Sequencing Project
 Direct Submission
 Submitted (29-MAR-2001) Genome Sequencing Center, Department of
 Genetics, Washington University School of Medicine, 4444 Forest
 Park Boulevard, St. Louis, MO 63108, USA
 COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
 computer analysis, using similarity information and the programs;
 GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
 GeneMark; <http://opal.biology.gatech.edu/GeneMark/>
 EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
 Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
 and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
 were kindly provided by Heladia Salgado, Julio Collado-Vides and
 Regunobu;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such

FEATURES

SOURCE

as compressions and repeats; all regions were covered by sequence from more than one ml3 subclone.

Location/Qualifiers

```
1..22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
/gene="phoH"
/feature="phoH"
/feature="synonym: STM1126"
434..439
/feature="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"
454..1308
/feature="phoH"
/note="similar to E. coli phoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=1
/product="phoB-dependent ATP-binding pho regulon component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/transl_table=1
/translation="MCANRQVENAGDVYLOKYLISMGKQAVIKARREARVLRSDS RSHQREBSYVTSVOMGVKVEALIGMARDSDTSVYKANEQAAYLNAIDSKOIFAT YGACGKTMVISAARAEALIHQDVERIIVTRPVLOADELGLPGDIEKPAFRRPV YDVLKRLGASFMOCYCLREIGKVEIAPFAYMRGRTPENAVIILDEAQNVAAMKMP LTRIGENTVIVNGDITQCCLPRGSRSLDALRFEDEVGVIVHPRKDCVASCALC ORTHAYS"
complement(1414..2303)
/gene="STM1127"
/complement(1414..2295)
/note="STM1127"
/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29% identity in aa 15 - 264"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/transl_table=1
/translation="MEPQPRLLKPKGLDITLGMOKSLTRASORIAQYIILAPPROVQ SSIALRETOGAEATVIRFCRTGYGKGFODPKDLIELATSSDDSPLLDAEVS SDDAHGLKLOKNTISNVLSETLLMDQVILGVVDLRLHCHSVYIFGVSSGITALD MKHLMRIGLRGDAVSNHMYMQATILKAGDAVGVSHSSTSPETVHSLRLAQACA TTVAITNLSGSPLEADFCILINRGQMLGDSIGTKAQLFVFDLLYTLVSSPE QARSKLRKTNALDMTK"
complement(2298..2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"
complement(2580..4088)
/gene="STM1128"
/complement(2580..4076)
/note="STM1128"
/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22% identity in aa 7 - 478"
/codon_start=1
/transl_table=1
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/transl_table=1
/translation="MITHSGEIVLVILFGYLLAMLVGVYSPROKADYPRGGGR VRGAAQSVAPATLSSITFMSIPAKAFSTWMTITIGQYLAIALPLVFFYIFPRK LKTSATVLEARDVRCRLPASMSPMLFHGRIATITFLVLAIRPFIADIVPLQA LISVWCIIYTMGGIEGVITWDVIGLLIGSALLIIVICLKVGGIIDEIFVTQDA DKFPATQFMSWTSTSPVLMIGLFLANIQFASQDVVQRYIVTSDISETKTLIT NAKLAVIIVFFAIGSALFYVYQGHQQLPAGNTGILPLFVTEMPIGIALIILA
```

```
AIFPAAGSSISSLSNISCFSNDIYORLSHKRRTPENRMKIAKLVILVAGLSSAAS
VILYMADESEIWDAPNSLIGMGSPMGLGIFPRKANGSAVLGIIIVITVYLG
RYATDLNPFYFGVIGLSLVISGVIFAPLFAFPAPPLTIDKEPEKVTL"
complement(4083..4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"
complement(4413..5099)
/gene="STM1129"
/complement(4413..5093)
/note="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/transl_table=1
/translation="MSILARLEOSVHENGGLIVSCOPVPGSPMDKPRIVAMAQAAAS AGAAYVIEGENTRTVRPHLSVYIIGIKRDLGSRVRIPIYQDDALAQACADIT AFDSFRRPVDISLIRIRLHGLMADSTVNEGISCHQKIEFGITLISYTG ITTVEPDLAVTQLSHAGCRVIAEGRVNTPALANALBHGMAVTGSAITRIHICQ WFSHAVKR"
complement(5094..5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"
5599..6759
/gene="STM1130"
5599..6759
/gene="STM1130"
/note="STM1130"
/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41% identity in aa 40 - 403"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/transl_table=1
/translation="WGMQKNPFKQMTLALCSVAITTSQVATLLPDIEPLKNGTG AINDNGVITVGLTACTGSKYKIDLKQHKOMERIKSPGARGQSVFPLDELIVFG GVGKSKESPLQVYSDVYKSPVNTQKDDTISPVGLTGTGKLBETWVLLTGVN EHFIDKFIIDIAAABADESEKKNVYVYFKPKADYFENKIVFIYNKENTWKNAGEL PGAGTGSSSVGMNNPFLMLINGELKPELPTDVIYRAMWDDNKLTLMLNSQLPSPRGQ QQBSIAGAPESGSHGUILVGGAGNFPKAKONTYNGKYSHEGIIKKRDEYGLINGH WQYGMKQKPLGVGSVSYGDEVFLIGENAKGPPSVVTSFTWRDNLILK"
6793..7497
/gene="STM1131"
6793..6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"
6805..7497
/gene="STM1131"
/note="similar to E. coli orf, hypothetical protein (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26% identity in aa 4 - 239"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein"
/protein_id="AAL20062.1"
/db_xref="GI:16419647"
/transl_table=1
/translation="MKINKYILGMVSIFARSSVLAQATLDRHEVADRTRINKORAI IEKLPNGIGPVYVAVKSGVDGEQDKHSLDVLVANAIELVSTRYKTKTDNPFVLOPGIT FESGPDSTIKPYLRQYINFDGVTMAGRYRYARNTKATYASDEKTRNPDYIYGVF DELKLENFPMWMSDQIKFDNKKTYNEHNVALMKLNSFTPYVEGVAVVRNNTDER QTRYRVLOQYHP"
7768..9060
/gene="STM1132"
7768..7773
/gene="STM1132"
/note="STM1132"
/note="putative RBS for STM1132; RegulonDB:STMS1H001403"
7780..9060
/gene="STM1132"
/note="similar to E. coli putative transport protein
```

(AAC77235.1): Blastp hit to AAC77235.1 (425 aa), 57% identity in aa 25 - 418"
 /codon_start=1
 /transl_table=11
 /product="putative sugar transport protein"

Query Match 73.0%; Score 332.8; DB 1; Length 22411;
 Best Local Similarity 83.1%; Pred. No. 6.2e-82;
 Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
 E-mail: parkhill@sanger.ac.uk
 Notes: Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/S_typhi/).

FEATURES

source location/Qualifiers

1..254050

/organism="Salmonella enterica subsp. enterica serovar Typhi"

/mol_type="genomic DNA"

/strain="CT18"

/db_xref="taxon:90370"

181..2793

/gene="STY1078"

/note="synonym: pepN"

181..2793

/gene="STY1078"

/note="Orthologue of E. coli pepN (AMPN ECOLI); Fasta hit to AMPN ECOLI (869 aa), 94% identity in 869 aa overlap"

/codon_start=1

/product="aminopeptidase N"

/protein_id="CAD08183.1"

/db_xref="GI:16502232"

/db_xref="GOA:082770"

/db_xref="SPTR:EMBL:082770"

/translation="MTQPOAKYRHDYRDPDYOTIDELFDLDAKTYVTVAISQAVR
 HGADPAPRLDGEDLTLSIHVNDAPMTAYAKEBGLITSDLEBRTLTAVELSPA
 NTALEGLYOSGDALCTCEAEGRHRTWYDREDYVLRFTTKITAKSKYPLLSVGN
 RVAOEELNENGRHWYQMODPPKPCYVPAIYAGPFDVLRFTTSSREVALBLVYDRG
 NLRAPAMENSTLAKSMKDETRGLRGLRDLDTWTVAVDFRNGAMENKELNTFNSKYV
 LARDTLDYDDYLDIERVETHEIFHWYTGRTVCRDMFQLSKELTVRDEFSDL
 GSRVNRISVNRTRGLQFAEDASPMAPTRPDKVLENNFYTLTYERGAETVIMTH
 TLDSERFQKMDLYFERHDSATCDDEPVOAMEASNDVLSHFRWVSQSGTPVTY
 KDYNPTEQETVLTISORTPATDAOEKPMLHPALIELYDNEGVTIPLOKGGHVNVA
 VLVNTQAEQTEFDENVYFQCPVPAALCEFSAPVLEKMSDDOLTTLPMHARNDSEW
 AAOGLATVYIKLVNVAHQOGPFLSPVHVAADFVLDKEDLPALAEIITLPSANE
 IAEFLFVYIDPIAIVARREALTRTLEADDEFILANVHLDYEVVDHGDIGKRLRN
 ACIRFLTFGEETLANTLSKOVARDANNITDALAAVAVAQLPORDYLMOEYDKRM
 ODGLVMDKMTILOSTSPARNVLETYRSLLKHSFMSNPNRIRSLRIFAFASNPAPFH
 AODSGOPVYENKLTIDLSNRNPOVASRLIEPLIRUKRYDDKQEKRAALQLNGLEN
 LSGDYLKIKTKALAA"

532..1332

/gene="STY1078"

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase family M1, score 245.10, E-value 1.9e-84"

1060..1089

/gene="STY1078"

/note="P800142 Neutral zinc metalloproteinases, zinc-binding region signature"

3001..4011

/gene="STY1079"

/note="synonym: pyrd"

3001..4011

/gene="STY1079"

/note="Orthologue of E. coli pyrd (PYRD ECOLI); Fasta hit to PYRD ECOLI (336 aa), 95% identity in 336 aa overlap"

/codon_start=1

/product="dihydroxycate dehydrogenase"

/protein_id="CAD08184.1"

/db_xref="GI:16502233"

/db_xref="GOA:082759"

/db_xref="SWISS-PROT:Q82759"

/translation="MYPPVFKALFQLEDERAHEFTFOQLRRITGTPLEALVROKVP
 KPYTKMGLTRNPLGLAAGLDKDCIDALMGESLEIGVTRPQDGNKPLRPL
 LVDAEGLINMGFNIGVNLVENYKKAHFDGILGINIGKNDITFVNGKDYILCME
 KVAIVAGYIAINISSPNTPLTLQYGDALDGLTLAIKKNKNDLVHKKYPAVAKI
 APDLCEBELIOVADSLRNIDGVLAINTTDLRSIVCGMKKCCQGTGLSGAPLDKST
 EITRRSLLEIKQLPIIGVGIDSVIAANEKIAAGAILVOYISGFIFKGPPLIKEIVT
 HI"

COMMENT

source

gene

CDS

misc_feature

misc_feature

gene

CDS

RESULT 14
 AL627269
 LOCUS 254050 bp DNA linear BCT 04-JUL-2003
 DEFINITION Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
 complete chromosome, segment 5/20.
 ACCESSION AL627269 AL513382
 VERSION AL627269.1 GI:16502231
 KEYWORDS
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (25-OCT-2001) Submitted on behalf of the Salmonella

```
misc_feature
3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHDHase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
misc_feature
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q82788"
/translation="MARIKPDNDNMYDEEDHDMMLDLANGMLFRSFRRLTPDA
FCPTGCVDDALVPSFEKCPDELTKORAEIVLNLVLAIRYKPKQMPKSHVVAH
GEMTPGCGDAASVLSPTABEQVNLIVPGENALCLLADPGVIAGRITQLDGAIK
IMNDRKPKQVHSHSFLSOAV"
/complement(4716..5825)
/gene="STY1081"
/complement(4716..5825)
/note="Orthologue of E. coli P75863; Faeta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82787"
/db_xref="SPTREMBL:Q82787"
/translation="WATLSRLFIHVKSMRGIGLTHALDISGLAFDRIFMVTESDGT
FITAROPOMVRFPSPLDGLHLPTRPGSSALVFPOTPDADPTELVNPAARVA
PTAINOMSGFSPSRDVOICWVGPOLTRPGRKNALVPLFAPAGVPLTLNPAASLDLO
RCPACVQMEQFRNLVSGVAAWEDSKVIRIGVIDVYKPCRCITFTVSPKGO
KHPGSEPLATQAFRTADNGVDVFGQVLIANSVIVAGVEVILATAPAKAAGATT
LDISTVPEKHPDVTIDWQGTFCGNNQVLLBQLENGIRIRIPYSCRAIGCGCRIR
LLEGVSPLKKSAMGDDSTIISCSVPFALALEN"
/complement(4737..4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical
protein h1016 H1016 SW:YCBY_HAEIN (P44524; P43945) Faeta
scores: E() : 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82786"
/db_xref="SPTREMBL:Q82786"
/translation="WNSLFPASTARGLEELIKLEKLGAVCCQVAVGSHVHOGDTRLI
YQSLMWSRLASRIILPMGECKVSDLDYFGVQALNMTWEINPGATFVHSGANDTI
RNSQYGMKVKDAIVDAFTRIKLLPREPNVDSPLDIRIVMLNKETASIALDLSGGLH
LRGYRDRGTGLAPIKETTLAAALVMSGMPGTPPLDPMCGSTLLIILAMAMVTDAPGL
```

```
Query Match 72.6%; Score 331.2; DB 1; Length 254050;
Best Local Similarity 82.9%; Pred. No. 1.7e-81;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

misc_feature
5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020,
E-value 3.7e-231"
E-value 3.7e-231"
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJUK_ECOLI (554 aa), 34% identity in 524 aa
overlap

1 ATGAAAAACAATTGTTATTTATGATGTTAACAATCTAGCGGCGCTGGGATTCGACC 60
88409 ATGAAAAACAATTGTTATTTATGATGTTAACAATCTAGCGGCGCTGGGATTCGACC 88468
61 GCAGAGGTTATGATTTAGTTAATTTAGTTAATTAATTAATTAATTAATTAATTAAT 120
88469 GCAGAGGTTATGATTTAGTTAATTTAGTTAATTAATTAATTAATTAATTAATTA 88528
121 TCTTCATTTAATGAGGAGCATTAATGTTGTCAGCTGGGAGTAAATATGATGCTCAGTTA 180
88529 TCTTCATTTAATGAGGAGCATTAATGTTGTCAGCTGGGAGTAAATATGATGCTCAGTTA 88588
181 CGGAGGAGGAGCTGAAAATTTTGGCGGTTTGGCGGAGGAGGAGGAGGAGGAGGAGG 240
88589 CGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 88648
241 AAGATTGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 300
88649 AAGATTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 88708
301 GATCCAGATATTTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
88709 GATCCAGATATTTGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 88768
361 GGTATTAAGCAATTTATTTACAGATAGTACTCAAAAAAGGCAATTTGATGTCAGAGA 420
88769 GGTATTAAGCAATTTATTTACAGATAGTACTCAAAAAAGGCAATTTGATGTCAGAGA 88828
421 CAGTCCGAATAGGCTATTTGCGGTGACACAGGTTAA 456
88829 CAGTCCGAATAGGCTATTTGCGGTGACACAGGTTAA 88864

RESULT 15
AE016840/c 301983 bp DNA linear BCT 21-MAR-2003
LOCUS Salmonella enterica subsp. enterica serovar Typh1 Ty2, section 7 of
DEFINITION 16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS Salmonella enterica subsp. enterica serovar Typh1 Ty2
```


THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaaacaaatcgtatc.....ctgcgtgacacacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*\n2: geneseqn1990s:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002s:*\n7: geneseqn2003as:*\n8: geneseqn2003bs:*\n9: geneseqn2003cs:*\n10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	3 AAC64620	AAC64620 Escherich
2	332.8	73.0	456	3 AAC64618	AAC64618 Salmonell
3	100	21.9	100	7 ACD68810	ACD68810 E. coli K
4	100	21.9	100	7 ACD68812	ACD68812 E. coli K
5	100	21.9	100	7 ACD68811	ACD68811 E. coli K
6	45.8	10.0	779	1 AAN40272	AAN40272 Sequence
7	38.2	8.4	4860	8 ADA31934	ADA31934 DNA encod
8	37.8	8.3	2000	7 ADA71938	ADA71938 Rice gene
9	37.4	8.2	456	3 AAC64623	AAC64623 Agfa::PT3
10	36.6	8.0	456	3 AAC64629	AAC64629 Agfa::PT3
11	35.2	7.7	4890	2 AAX34648	AAX34648 Promoter
12	35	7.7	646	2 AAO62647	AAO62647 Fibronect
13	34.6	7.6	361	2 AAQ73066	AAQ73066 Agfa sequ
14	34.6	7.6	361	2 AAT74141	AAT74141 Salmonell
15	34.6	7.6	456	2 AAQ87467	AAQ87467 Agfa sequ
16	34.6	7.6	456	2 AAT74142	AAT74142 Salmonell
17	34.6	7.6	456	3 AAC64622	AAC64622 Agfa::PT3
18	34.6	7.6	456	3 AAC64617	AAC64617 Salmonell
19	34.6	7.6	456	3 AAC64625	AAC64625 Agfa::PT3
20	34.6	7.6	456	3 AAC64624	AAC64624 Agfa::PT3
21	34.4	7.5	4455	7 ABT17054	ABT17054 Human MP2
22	34.4	7.5	5184	9 ADE25651	ADE25651 Human CDN
23	34.4	7.5	6325	4 AAK80005	AAK80005 Human imm

C	24	34.4	7.5	9555	7 ABX34542	ABX34542 Human mdd
	25	34.2	7.5	1083	5 AAS76745	AAS76745 DNA encod
	26	33.4	7.5	12426	4 AAS64682	AAS64682 Tumour su
	27	33.8	7.4	1149	3 AAB65947	AAB65947 E. coli p
	28	33.8	7.4	1149	7 ACA18571	ACA18571 Prokaryot
	29	33.8	7.4	2655	6 ABN70427	ABN70427 Streptoco
	30	33.8	7.4	2655	6 ABN67963	ABN67963 Streptoco
C	31	33.8	7.4	6108	4 ABL08478	ABL08478 Drosophil
	32	33.8	7.4	6808	4 ABL08488	ABL08488 Drosophil
	33	33.8	7.4	110000	6 ABN71527_06	Continuation (7 of
	34	33.6	7.4	606	3 AAC43625	AAC43625 Arabidops
	35	33.6	7.4	1299	7 ACA48239	ACA48239 Prokaryot
	36	33.2	7.3	1371	7 ACA29389	ACA29389 Prokaryot
	37	33.2	7.3	4590	5 AAB24065	AAB24065 Yeast AOD
	38	33	7.2	1269	7 ACA48121	ACA48121 Prokaryot
	39	33	7.2	1837	7 ACA47731	ACA47731 Prokaryot
C	40	33	7.2	110000	6 ABO69245_18	Continuation (19 o
	41	33	7.2	319630	6 ABO67194	ABQ67194 Listeria
	42	32.6	7.1	351	7 ABX54413	ABX54413 Bovine ES
	43	32.6	7.1	967	5 AAS74657	AAS74657 DNA encod
C	44	32.6	7.1	1476	6 AAS97253	AAS97253 Neisseria
	45	32.6	7.1	1476	6 AAS97288	AAS97288 Neisseria

ALIGNMENTS

RESULT 1
AAC64620
ID AAC64620 standard; DNA; 456 BP.
XX
AC AAC64620;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgB DNA sequence SEQ ID NO:4.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
XX (UTVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW,
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36344.
XX
PS The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgB and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

SO Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 3.3e-130; Mismatches 0; Indels 0; Gaps 0;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAAGC 60
 DB 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAAGC 60
 QY 61 GCAGCAGGTTATGATTTAGTATTTACGATTTAATCTTGGCGCTGGGATTGCAAGC 60
 DB 61 GCAGCAGGTTATGATTTAGTATTTACGATTTAATCTTGGCGCTGGGATTGCAAGC 60
 QY 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 120
 DB 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 120
 QY 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
 DB 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
 QY 181 CGCAGGAGGAGGCTCAAACTTTGGCGCTGGGCGCAAGAGTATGAGCAGCAGCAGC 240
 DB 181 CGCAGGAGGAGGCTCAAACTTTGGCGCTGGGCGCAAGAGTATGAGCAGCAGCAGC 240
 QY 241 AAGATTGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 300
 DB 241 AAGATTGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 300
 QY 301 GATCCAGATTTTCGCAAGGCTTATGATTTATCTGCGATGATTTATCCGAAAGGTTCT 360
 DB 301 GATCCAGATTTTCGCAAGGCTTATGATTTATCTGCGATGATTTATCCGAAAGGTTCT 360
 QY 361 GGTAAATTAAGCAATTTTACAGATGATGTACTCAAAAACGCAATTGATGCGAGAGA 420
 DB 361 GGTAAATTAAGCAATTTTACAGATGATGTACTCAAAAACGCAATTGATGCGAGAGA 420
 QY 421 CATGCCAATGCTATTTGCGTGACACAGCTTAA 456
 DB 421 CATGCCAATGCTATTTGCGTGACACAGCTTAA 456
 QY 421 CATGCCAATGCTATTTGCGTGACACAGCTTAA 456
 DB 421 CATGCCAATGCTATTTGCGTGACACAGCTTAA 456

RESULT 2
 AAC64618
 AAC64618 standard; DNA; 456 BP.

AAC64618;
 26-FEB-2001 (first entry)
 DE *Salmonella enteritidis* AgfB DNA sequence SEQ ID NO:2.
 KW *Salmonella*; agfA; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS *Salmonella enteritidis*.
 XX
 PN WO200060102-A2.

PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000MO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (YVVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay MW,
 PI WPI; 2000-672631/65.
 DR P-PSDB; AAB36342.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 134; 139p; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TFP) nucleation depended
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

SO Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;

Query Match 73.0%; Score 332.8; DB 3; Length 456;
 Best Local Similarity 83.1%; Pred. No. 3.1e-92; Mismatches 77; Indels 0; Gaps 0;

Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
 QY 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAAGC 60
 DB 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAAGC 60
 QY 61 GCAGCAGGTTATGATTTAGTATTTACGATTTAATCTTGGCGCTGGGATTGCAAGC 120
 DB 61 GCAGCAGGTTATGATTTAGTATTTACGATTTAATCTTGGCGCTGGGATTGCAAGC 120
 QY 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
 DB 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
 QY 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240
 DB 121 TCTTCATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240
 QY 181 CGCAGGAGGAGGCTCAAACTTTGGCGCTGGGCGCAAGAGTATGAGCAGCAGCAGC 240
 DB 181 CGCAGGAGGAGGCTCAAACTTTGGCGCTGGGCGCAAGAGTATGAGCAGCAGCAGC 240
 QY 241 AAGATTGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 300
 DB 241 AAGATTGACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 300
 QY 301 GATCCAGATTTTCGCAAGGCTTATGATTTATCTGCGATGATTTATCCGAAAGGTTCT 360
 DB 301 GATCCAGATTTTCGCAAGGCTTATGATTTATCTGCGATGATTTATCCGAAAGGTTCT 360

Db 301 GATGCACTATATGCAAGCGCTTACGCTATAGTGCAGCTATTATTCAGAAAGCTTCT 360
Oy 361 GGTATATAAGCAATATTATACACAGTATGCTCTAAAAAGCGCAATTGTAGTCAGAGA 420
Db 361 GGAATATAAGCCCAATTATACCGTACCGTACGCGAAGAAACAGAGTGTAGTCAGAAA 420
Oy 421 CAGTCGCAATATGCGCTATGCGGTGACACAGCTTAA 456
Db 421 CAGTCGCTATGCTATTGCGCTCACCCAGCTTAA 456

RESULT 3
ACD68810
ID ACD68810 standard; DNA; 100 BP.
ACD68810;
ACD68810;
18-SEP-2003 (first entry)
Dt 18-SEP-2003 (first entry)
Df E. coli K12 MG1655 biochip probe SEQ ID 80.
Df Biochip; gene expression; gut; diagnostic; detection; probe; ss.
Km Escherichia coli.
Os Epi260592-A1.
Pn Epi260592-A1.
Pd 27-NOV-2002.
Pf 17-MAY-2001; 2001EP-00112179.
Pr 17-MAY-2001; 2001EP-00112179.
Pa (MMGB-) MWG-BIOTECH AG.
Pi Donner H, Drescher B, Huber A, Weber J;
Dr WPI; 2003-241155/24.
Pt Biochip containing probes complementary with open reading frames in
Pt Escherichia coli K12, useful for detecting gene expression and expression
Pt patterns.
Ps Claim 3; Page 23; 2004pp; German.
Xx This invention describes a novel biochip comprising probe spots, each
Cc containing many identical probes. The probes are nucleotide sequences of
Cc 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
Cc least one includes a segment of at least 20 bases identical with, or
Cc complementary to, a segment of an open reading frame (orf) of Escherichia
Cc coli K12. The biochip is used for specific detection of gene expression
Cc in K12 and for determining the gene expression pattern, e.g. for
Cc diagnostic determination of which E. coli strains are present in the gut,
Cc and to determine the effects of e.g. growth media on gene expression. The
Cc biochip provides as comprehensive as possible detection of the K12
Cc genome, with simultaneous analysis of many different genes with a single
Cc device, and comparison of gene expression between K12 and its mutants or
Cc other E. coli strains in a single experiment. Apart from qualitative and
Cc quantitative information about gene expression, it also allows
Cc measurements of population densities for the various strains. The use of
Cc synthetic oligonucleotides for preparation of probes allows free
Cc variation in probe length and ensures high purity (and thus selectivity,
Cc reactivity and reproducibility); also synthetic probes are generally
Cc shorter than probes prepared by polymerase chain reaction. ACD68731 to
Cc ACD81540 represent oligonucleotide probes used with the biochip described
Cc in the invention

Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;
Query Match 21.9%; Score 100; DB 7; Length 100;
Best Local Similarity 100.0%; Pred. No. 9.5e-21;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 266 ACCTTGATATATGATCAGCGCGGCGACGTGCAACGATGCCAGTATTTCCGAAGTCTT 325
Db 1 ACCTTGATATATGATCAGCGCGGCGACGTGCAACGATGCCAGTATTTCCGAAGTCTT 60
Oy 326 ATGGTATATACGCGATATATTCAGAAAGTCTGTGTA 365
Db 61 ATGGTATATACGCGATATATTCAGAAAGTCTGTGTA 100

RESULT 4
ACD68812
ID ACD68812 standard; DNA; 100 BP.
ACD68812;
ACD68812;
18-SEP-2003 (first entry)
Dt 18-SEP-2003 (first entry)
Df E. coli K12 MG1655 biochip probe SEQ ID 82.
Df Biochip; gene expression; gut; diagnostic; detection; probe; ss.
Km Escherichia coli.
Os Epi260592-A1.
Pn Epi260592-A1.
Pd 27-NOV-2002.
Pf 17-MAY-2001; 2001EP-00112179.
Pr 17-MAY-2001; 2001EP-00112179.
Pa (MMGB-) MWG-BIOTECH AG.
Pi Donner H, Drescher B, Huber A, Weber J;
Dr WPI; 2003-241155/24.
Pt Biochip containing probes complementary with open reading frames in
Pt Escherichia coli K12, useful for detecting gene expression and expression
Pt patterns.
Ps Claim 3; Page 23; 2004pp; German.
Xx This invention describes a novel biochip comprising probe spots, each
Cc containing many identical probes. The probes are nucleotide sequences of
Cc 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
Cc least one includes a segment of at least 20 bases identical with, or
Cc complementary to, a segment of an open reading frame (orf) of Escherichia
Cc coli K12. The biochip is used for specific detection of gene expression
Cc in K12 and for determining the gene expression pattern, e.g. for
Cc diagnostic determination of which E. coli strains are present in the gut,
Cc and to determine the effects of e.g. growth media on gene expression. The
Cc biochip provides as comprehensive as possible detection of the K12
Cc genome, with simultaneous analysis of many different genes with a single
Cc device, and comparison of gene expression between K12 and its mutants or
Cc other E. coli strains in a single experiment. Apart from qualitative and
Cc quantitative information about gene expression, it also allows
Cc measurements of population densities for the various strains. The use of
Cc synthetic oligonucleotides for preparation of probes allows free
Cc variation in probe length and ensures high purity (and thus selectivity,
Cc reactivity and reproducibility); also synthetic probes are generally
Cc shorter than probes prepared by polymerase chain reaction. ACD68731 to
Cc ACD81540 represent oligonucleotide probes used with the biochip described
Cc in the invention

Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;
Query Match 21.9%; Score 100; DB 7; Length 100;
Best Local Similarity 100.0%; Pred. No. 9.5e-21;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 123 TTCATTATATCAGCAGCATATATGCTCAAGTGGAGCTAATATATATGCTCAGTTAGC 182

Db 1 TTTATTTATATAGGACGACCAATATGTGACGTGGACCTATATAGTCTCATGTTACG 60
 QY 183 GCAGGAGGCTCAAACTTTGGCGGTGTGTGGCAAGAA 222
 Db 61 GCAGGAGGCTCAAACTTTGGCGGTGTGTGGCAAGAA 100

RESULT 5

ACD68811
 ID ACD68811 standard; DNA; 100 BP.
 AC ACD68811;
 XX
 DT 18-SEP-2003 (first entry)
 XX
 DE E. coli K12 MG1655 biochip probe SEQ ID 81.
 XX
 KW Biochip; gene expression; gut; diagnostic; detection; probe; ss.
 XX
 OS Escherichia coli.
 XX
 PN EP1260592-A1.
 XX
 PD 27-NOV-2002.
 XX
 PE 17-MAY-2001; 2001EP-00112179.
 XX
 PR 17-MAY-2001; 2001EP-00112179.
 XX
 PA (MWGB-) MWG-BIOTECH AG.
 XX
 PI Donner H, Dreacher B, Huber A, Weber J;
 XX
 DR WPI; 2003-241155/24.
 XX
 PT Biochip containing probes complementary with open reading frames in
 PT Escherichia coli K12, useful for detecting gene expression and expression
 PT patterns.
 XX
 PS Claim 3; Page 23; 2004pp; German.

CC This invention describes a novel biochip comprising probe spots, each
 CC containing many identical probes. The probes are nucleotide sequences of
 CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
 CC least one includes a segment of at least 20 bases identical with or
 CC complementary to, a segment of an open reading frame (orf) of Escherichia
 CC coli K12. The biochip is used for specific detection of gene expression
 CC in K12 and for determining the gene expression pattern, e.g. for
 CC diagnostic determination of which E. coli strains are present in the gut,
 CC and to determine the effects of e.g. growth media on gene expression. The
 CC biochip provides as comprehensive analysis of many different genes in the
 CC genome, with simultaneous analysis as possible detection of the K12
 CC device, and comparison of gene expression between K12 and its mutants or
 CC other E. coli strains in a single experiment. Apart from qualitative and
 CC quantitative information about gene expression, it also allows
 CC measurements of population densities for the various strains. The use of
 CC synthetic oligonucleotides for preparation of probes allows free
 CC variation in probe length and ensures high purity (and thus selectivity,
 CC reactivity and reproducibility); also synthetic probes are generally
 CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
 CC ACD61540 represent oligonucleotide probes used with the biochip described
 CC in the invention
 XX

SO Sequence 100 BP; 31 A; 20 C; 29 G; 20 T; 0 U; 0 Other;

Query Match 21.9%; Score 100; DB 7; Length 100;
 Best Local Similarity 100.0%; Pred. No. 9.5e-21;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GTTGTGCGCCAGAAAGTAGTAGCAACCGGCAAGATTACGACAGAGAGATTATAC 267
 Db 1 GTTGTGCGCCAGAAAGTAGTAGCAACCGGCAAGATTACGACAGAGAGATTATAC 60

QY 268 CTTCATATATATGATCAGCGCGGACAGTGCACCAACCATGCCA 307
 Db 61 CTTCATATATATGATCAGCGCGGACAGTGCACCAACCATGCCA 100

RESULT 6

AAN40272
 ID AAN40272 standard; CDNA; 779 BP.
 XX
 AC AAN40272;
 XX
 DT 09-JAN-1992 (first entry)
 XX
 DE Sequence of a clone numbered as Ag16 which corresponds to the S antigen
 DE of isolate FC27.
 XX
 KW Malaria; vaccine; antigen; epitope; immune response; ss.
 XX
 OS Plasmodium falciparum.

FT Key Location/Qualifiers
 FT CDS 1..777
 FT misc_feature 1..8
 FT repeat_region 10..768
 FT /tag= a
 FT /label= linker
 FT /tag= c
 FT /rpt unit= 10..42
 FT /note= "final base of certain codons may differ but AA SQ
 FT 772..779
 FT encoded stays the same"
 FT /tag= b
 FT /label= linker

PN WO8402917-A.
 XX
 PD 02-AUG-1984.
 XX
 PF 27-JAN-1984; 84WO-AU000016.
 XX
 PR 28-JAN-1983; 83AU-00007843.
 PR 10-JUN-1983; 83AU-00009788.
 PR 26-JAN-1984; 84AU-00023842.
 XX
 PA (HALI-) HALI INST MED RES.
 PA (KEMP/) KEMP D J.
 PA (AUME-) AUST INST MED RES.
 PA (SARA-) SARAMANE PTY LTD.

PI Kemp DJ, Anders RF, Coppel RL, Brown G, Saint R, Cowman AF;
 PI Mitchell GF;
 DR WPI; 1984-201410/32.
 DR P-PSDB; AAN40376.
 XX
 PT Expression of plasmodium falciparum poly(peptide(s) from CDNA - for use
 PT in immunisation against malaria infection.
 XX
 PS Claim 5; Fig 1; 62pp; English.

CC The inventors claim a DNA SQ which comprises a nucleotide sequence
 CC corresp. to all or part of Plasmodium falciparum RNA. Also claimed is a
 CC synthetic peptide or polypeptide displaying the antigenicity of all or
 CC part of a P.falciparum antigen. The base sequence of clone Ag16 (see
 CC AAN40272) indicates that the antigen coded for by this partial
 CC polynucleotide sequence has a homologous repeat structure of 11 AAs
 CC tandemly repeated 23 times
 XX

SO Sequence 779 BP; 294 A; 125 C; 217 G; 143 T; 0 U; 0 Other;

Query Match 10.0%; Score 45.8; DB 1; Length 779;

Best Local Similarity 47.7%; Pred. No. 0.0011;
Matches 134; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy 126 ATTTAATCAGGCGACCATTAATGCTCAATATATATAGTCTCAGTTACGGCA 185
Db 198 ATTGAAGATCTCTCAAAAGCTAGTCAAGTATGATCCCTGCAAAAGCAAGTCA 257
Qy 186 GGGAGGCTCAAACTTTTGGCGGTGTTGGCGAAGAGTAGTAGCAACCGGCAAAAT 245
Db 258 AGGAGGATTAGAGATCTCTGCAAAAGCTAGTCAAGTATGATCCCTGCAAAAGC 317
Qy 246 TGACCAAGACAGAGATTATTAACCTTGATATATGATCAGCGGCGAGTGCACGATGC 305
Db 318 TAGTCAGAGAGATTAGAGATCTCTGCAAAAGCTAGTCAAGTATGATGAAAGTCTGC 377
Qy 306 CAGTATTTCCGAGGTCTTATGATATCTGCGATGATTATCCAGAAAGTTCTGTTAA 365
Db 378 AAAAGCTAGTCAAGTGTGATTAGAGATCCCGCAAGGCTAGTCAAGGTGATTAGAGA 437
Qy 366 TAAAGCAAAATTTACACAGTATGCTCAAAAACGGCAA 406
Db 438 TCTGCAAAAGCTAGTCAAGTGTGATTAGAGATCCGCAA 478

RESULT 7

ADA1934
ID ADA1934 standard; DNA; 4860 BP.

AC ADA1934;

DT 20-NOV-2003 (first entry)

DE DNA encoding Acinetobacter baumannii protein #3221.

KW ds; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KM vaccine; plant biocontrol agent.

OS Acinetobacter baumannii.

PN US6562958-B1.

PD 13-MAY-2003.

PF 04-JUN-1999; 99US-00328352.

PR 09-JUN-1998; 98US-0088701P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Breton G, Bush D;

DR WPI; 2003-576092/54.

DR P-PSDB; ADA36060.

PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.

PS Example; SEQ ID NO 3221; 328bp; English.

CC The invention relates to isolated Acinetobacter baumannii nucleic acids.

CC The A. baumannii nucleic acids and polypeptides are useful as reagents

CC for diagnosing a bacterial disease, as components of antibacterial

CC vaccines, as targets for antibacterial drugs, to detect the presence of

CC A. baumannii and other Acinetobacter species in a sample, in screening

CC compounds for the ability to interfere with the A. baumannii life cycle

CC or to inhibit A. baumannii infection, and as biocontrol agents for

CC plants. The present sequence represents DNA encoding an A. baumannii

CC protein.

CC Sequence 4860 BP; 1639 A; 752 C; 998 G; 1471 T; 0 U; 0 Other;

Query Match 8.4%; Score 38.2; DB 8; Length 4860;

Best Local Similarity 45.9%; Pred. No. 0.51; Indels 0; Gaps 0;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

Qy 106 AATGAATGAGTAAGTCTTCAATTAATCAGGCGACCATTAATGCTCAAGTGGACCTAAT 165
Db 3445 AATACTATTTGGTAAATATTATTAATGATGATGATGATGATGATGATGATGATGATGAT 3504
Qy 166 AATAGTCTCAGTTACGCGAGGAGGCTCAAAACTTTTGGCGGTGTTGGCGGCAAGAGT 225
Db 3505 ATTTCACCTATTTATGTTGGTGGAGGCAATGATCGATCAAGGTGGTGCAGAAATGAT 3564
Qy 226 AGTAGCAACCGGCGCAAGATTGACAGACAGAGATTATTAACCTTGATATATGATGATGAT 285
Db 3665 TATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3624
Qy 286 GCGGCGAGTGCACGATGCGAGTATTTCCGAGGTCTTATGATGATGATGATGATGATGAT 345
Db 3625 TATGTTGTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3684
Qy 346 ATCCAGAAAGGTTCTGTTATTAAGCAATATTACACAGTATG 388
Db 3685 TTACAGAGTCAAGCTGGGAATGACACTTATTTATGATTAAG 3727

RESULT 8

ADA1938/c
ID ADA1938 standard; DNA; 2000 BP.

AC ADA1938;

DT 20-NOV-2003 (first entry)

DE Rice gene, SEQ ID 5263.

KW Plant; bacterial infection; fungal infection; viral infection; rice;

KM gene; ds.

OS Oryza sativa.

PN WO2003000898-A1.

PD 03-JAN-2003.

PF 22-JUN-2001; 2001WO-IB001105.

PR 22-JUN-2001; 2001WO-IB001105.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y;

PI Katsagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

PI WPI; 2003-175290/17.

PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.

PS Claim 27; SEQ ID NO 5263; 899bp; English.

CC The present invention relates to a method (M1) for identifying genes

CC involved in plant resistance or response to pathogenic infection. M1

CC comprises identifying a gene whose expression is significantly altered in

CC the incompatible interaction of plant gene expression relative to

CC expression of the gene in an uninfected plant, in a mutant plant that

CC or in a corresponding incompatible or compatible interaction. (M1) is

CC useful for conferring resistance to resistance or tolerance to a plant to

CC bacterial, fungal or viral infection. The present sequence was used to

CC illustrate the invention.

PR 03-NOV-1992; 92US-00970846.
 XX (REAT/) REA T S.
 PA Normark S, Olsen A;
 XX WPI, 1994-167484/20.
 DR P-PSDB; AAR52664.
 XX New fibronectin binding protein curlin from *E. coli* - used for the
 PT treatment of infections caused by *E. coli* and for immunisation against *E.*
 PT *coli* infection.
 XX Disclosure; Fig 2; 33pp; English.
 PS
 CC The *fmbA* gene of *E. coli* cattle fecal isolate A012 includes DNA encoding
 CC a new fibronectin binding protein (fBP), curlin, which is produced as
 CC curli pili when expressed in *E. coli* (pFmb20) DSM 4585. The *fmbA* gene
 CC sequence from pFmb20 is given in AA062647, and includes the sequence
 CC encoding the 17 kDa curlin subunit (amino acids 21-142 of AAR52664).
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX
 SQ Sequence 646 BP; 168 A; 146 C; 161 G; 171 T; 0 U; 0 Other;
 Query Match 7.7%; Score 35; DB 2; Length 646;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 416 AGAGACAGTCGCAATGCGCTATTTCGGGTACACAA 450
 Db 1 AGAGACAGTCGCAATGCGCTATTTCGGGTACACAA 35
 RESULT 13
 ID AAQ73066 standard; DNA; 361 BP.
 XX
 AC AAQ73066;
 XX
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KW Salmonella, Agfa; vaccine, genetic immunization; ds.
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..359
 FT /*tag= a
 FT /note= "Agfa"
 FT 37..60
 FT /*tag= d
 FT /note= "TAF5 primer (pair with TAF6)"
 FT 52..69
 FT /*tag= b
 FT /note= "TAF3 primer (pair with TAF4)"
 FT complement(103..129)
 FT /*tag= e
 FT /note= "TAF6 primer (pair with TAF5)"
 FT complement(292..402)
 FT /*tag= c
 FT /note= "TAF4 primer (pair with TAF3)"
 PN W09425598-A2.
 PD 10-NOV-1994.
 XX
 PF 26-APR-1994; 94WO-IB000207.
 XX
 PR 26-APR-1993; 93US-00054452.

XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX (KING/) KING J.
 XX
 XX Kay WW, Collinson SK, Cloughier SC, Doran JL;
 XX WPI, 1994-358275/44.
 DR P-PSDB; AAR62761.
 XX
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7A; 95pp; English.
 CC
 CC The DNA encodes the Salmonella enteritidis27655-3b *trpH*A mutant strain
 CC *agfa* gene cloned into pUC19. The DNA and isolated proteins are used in
 CC genetic immunization and vaccine compositions, respectively, to elicit an
 CC immune response to Salmonella in animals (e.g. food producing animals)
 CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
 CC AUG-2003 to correct OS field.)
 CC
 XX
 SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
 Query Match 7.6%; Score 34.6; DB 2; Length 361;
 Best Local Similarity 53.3%; Pred. No. 2.3;
 Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 QY 277 ATTGATCAGCGCGGACATGCCAGATGCTTTGCAAGGTGCTTATGTAATACT 336
 Db 145 ATTACCCAGAGCGGCTTATGTTAGTAACGCCCGCATGTAGGCCAGGGTGCGATTAATAGTACT 204
 QY 337 GCGATGATTAATCCAGAAAGTTCTGTATATTAAGCAATATTACAGTATAGTACTCAA 396
 Db 205 ATTGAACGTACTCAAAAGGTTTCAGAAATATATGCCACATGCACGTGAGACGCTAAA 264
 QY 397 AAAACGCGCAATTGACT 413
 Db 265 AACTCCGATATTACTGT 281
 RESULT 14
 ID AAT74141
 XX
 ID AAT74141 standard; DNA; 361 BP.
 XX
 AC AAT74141;
 XX
 DT 25-MAR-2003 (revised)
 DT 30-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b *trpH*A mutant *agfa* gene fragment.
 XX
 KW Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody; ds.
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..360
 FT /*tag= a
 FT /label= *agfa_gene_fragment*
 FT 16..60
 FT /*tag= b
 FT /label= *primer_TAF5*
 FT 52..69
 FT /*tag= c
 FT /label= *primer_TAF3*
 FT complement(103..128)
 FT /*tag= d
 FT /label= *primer_TAF6*
 FT complement(294..312)
 FT /*tag= e
 FT /label= *primer_TAF4*
 PN US5635617-A.

XX 03-JUN-1997.
 PD 26-APR-1994; 94US-00233788.
 XX 26-APR-1993; 93US-00054452.
 XX (UVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (UVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;
 DR WPI; 1997-309886/28.
 DR P-PSDB; AAM23569.
 XX
 PT Isolated *Salmonella* gene *agfa* - used for diagnosis of *Salmonella* or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX
 PS Claim 1; Col 107-110; 85pp; English.
 XX
 CC The present sequence represents an isolated *agfa* gene fragment derived
 CC from *Salmonella* enteritidis 27655-3b Typhoid mutant strain. The nucleic
 CC acid can be used to provide diagnostic assays for *Salmonella* and/or
 CC enteropathogenic bacteria of the family Enterobacteria. It can also be
 CC used to provide proteins and antibodies which can be used for assays. The
 CC nucleic acid sequence can be used to provide probes or primers which can
 CC specifically hybridise to nucleic acid molecules from greater than 99% of
 CC *Salmonella* strains that are pathogenic to warm-blooded animals relative
 CC to nucleic acid molecules from virtually all other microbial organisms.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 CC
 XX
 SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 7.6%; Score 34.6; DB 2; Length 361;
 Best Local Similarity 53.3%; Pred. No. 2.3;
 Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 QY 277 ATTGATCAGCGGCGAGTCCAGATGCGGATTTTCGCAAGGCTTATGTAATCT 336
 DB 145 ATTACCCAGAGCGGTTATGTAACGCCCGCATGTAGCCGAGGTGCGATTAATGTA 204
 QY 337 GCGATGATTAATCCAGAAAGTTCTGTGTAATAAGCAATATTACACAGTATGTA 396
 DB 205 ATGAACTGACTCAGAAATGTTTCAGAAATATATGCAACATGCACAGTGAACGCTAAA 264
 QY 397 AAAACGGCAATTGTA 413
 DB 265 AACTCCGATATTACTGT 281

RESULT 15
 ID AAQ87467 standard; DNA; 456 BP.
 XX
 AC AAQ87467;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE *Agfa* sequence.
 XX
 KM *Salmonella*; *Agfa*; vaccine; genetic immunization; *de*.
 KM
 OS *Salmonella*.
 XX
 XX Key Location/Qualifiers
 FH 1.454
 FT CDS /tag= a
 FT /note= "Agfa"
 XX
 PN W09425598-A2.
 XX
 PD 10-NOV-1994.
 XX

PF 26-APR-1994; 94WO-IB000207.
 XX 26-APR-1993; 93US-00054452.
 XX (UVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 PI Kay WW, Collinson SK, Clouchier SC, Doran JL;
 DR WPI; 1994-358275/44.
 DR P-PSDB; AAR74625.
 XX
 PT Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compens. contg. fibrial type proteins.
 XX
 PS Disclosure; Fig 7B; 95pp; English.
 XX
 CC The DNA encodes the *Salmonella* *Agfa* protein. The DNA and isolated
 CC proteins are used in genetic immunization and vaccine compositions,
 CC respectively, to elicit an immune response to *Salmonella* in animals (e.g.
 CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 CC
 XX
 SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 7.6%; Score 34.6; DB 2; Length 456;
 Best Local Similarity 53.3%; Pred. No. 2.5;
 Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 QY 277 ATTGATCAGCGGCGAGTCCAGATGCGGATTTTCGCAAGGCTTATGTAATCT 336
 DB 208 ATTACCCAGAGCGGTTATGTAACGCCCGCATGTAGCCGAGGTGCGATTAATGTA 267
 QY 337 GCGATGATTAATCCAGAAAGTTCTGTGTAATAAGCAATATTACACAGTATGTA 396
 DB 268 ATGAACTGACTCAGAAATGTTTCAGAAATATATGCAACATGCACAGTGAACGCTAAA 327
 QY 397 AAAACGGCAATTGTA 413
 DB 328 AACTCCGATATTACTGT 344

Search completed: March 15, 2004, 17:51:50
 Job time : 248.684 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaataacaatcgtatc.....ctcgctgcacacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	39.6	8.7	7218	1	US-08-232-463-14
2	38.2	8.4	4860	4	US-09-328-352-3221
3	34.6	7.6	361	1	US-08-233-788A-56
4	34.6	7.6	456	1	US-08-233-788A-58
5	31.8	7.0	580073	4	US-08-545-528D-1
6	31.8	7.0	640681	4	US-08-790-988-1
7	31.4	6.9	525	4	US-09-107-532A-2325
8	31.4	6.9	2785	4	US-08-956-171E-133
9	31.4	6.9	53332	4	US-09-801-861-3
10	31.4	6.8	62902	4	US-09-596-002-32
11	31.2	6.8	606	4	US-09-540-236-1518
12	31.2	6.8	1404	4	US-09-134-001C-398
13	31	6.8	2139	4	US-09-059-584-50
14	31	6.8	1664976	4	US-08-916-421B-1
15	30.6	6.7	832	4	US-09-621-976-3813
16	30.6	6.7	1664976	4	US-08-916-421B-1
17	30.4	6.7	831	4	US-09-328-352-3901
18	30.4	6.7	1023	4	US-09-540-236-1556
19	30.4	6.7	1622	4	US-09-107-532A-2544
20	30.4	6.7	1622	4	US-08-956-171E-732
21	30.4	6.7	1776	4	US-09-134-000C-3324
22	30.4	6.7	1879	4	US-09-023-655-661
23	30.4	6.7	2001	3	US-08-714-918-82
24	30.4	6.7	2001	3	US-09-265-315-92
25	30.4	6.7	2001	3	US-09-265-315-92
26	30.4	6.7	2001	3	US-09-266-417-92
27	30.4	6.7	2001	4	US-09-528-709-92

ALIGNMENTS

28	30.4	6.7	2001	4	US-09-527-745-92	Sequence 92, Appl
29	30.4	6.7	99629	4	US-09-596-002-37	Sequence 37, Appl
C 30	30.2	6.6	417	4	US-09-621-976-10111	Sequence 10111, A
31	30.2	6.6	2364	3	US-09-172-045-1	Sequence 1, Appl1
32	30.2	6.6	2364	4	US-09-342-325C-1	Sequence 1, Appl1
33	30.2	6.6	4215	4	US-09-620-312D-295	Sequence 295, App
34	30.2	6.6	13508	4	US-08-956-171E-120	Sequence 120, App
35	30.2	6.6	152321	3	US-09-128-155-16	Sequence 16, Appl
36	30.2	6.6	176373	3	US-09-128-155-17	Sequence 17, Appl
37	30	6.6	2136	2	US-08-949-941B-1	Sequence 1, Appl1
C 38	29.8	6.5	7091	3	US-08-975-762-46	Sequence 46, Appl
C 39	29.8	6.5	7091	3	US-09-295-028-46	Sequence 46, Appl
C 40	29.8	6.5	7091	4	US-09-106-582-46	Sequence 46, Appl
C 41	29.8	6.5	7091	4	US-09-159-469-46	Sequence 46, Appl
C 42	29.8	6.5	7091	4	US-09-693-542-46	Sequence 46, Appl
C 43	29.8	6.5	13158	2	US-08-687-080-105	Sequence 105, App
C 44	29.4	6.4	703	4	US-09-016-434-120	Sequence 120, App
C 45	29.2	6.4	369	4	US-09-328-352-3808	Sequence 3808, App

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pC-Fls
US-08-232-463-14
Query Match 8.7%; Score 39.6; DB 1; Length 7218;

```

RESULT 2
US-09-328-352-3221
; Sequence 3221, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANN11 FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
; LENGTH: 4660
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3221

```

RESULT 3
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617

US-08-233-788A-56
Sequence 56, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/33,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043, 403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-788A-56

Query Match	Best Local Similarity	Score	DB 1	Length
Matches 73; Conservative	53.3%;	Pred. No. 0.087;	Mismatches 64; Indels 0; Gaps 0;	
QY	277	ATTGATCAGCGCGGACGTCGCAACGATGTCAGTATTTGGCAGGCTTATGTA	ACT	336
DB	145	ATTACCCAGCGGTTATGTGTAAAGCGCGCATGTAGCCAGGTCGGATATATG	ACT	304
QY	337	GGCAGTATTAACCAAGAGTTTCGTGAATTAAGCAAAATTTACACAGTAGTACT	CAA	396
DB	205	AATGACAGACTCAGAAATGTTTCAGAAATTAATGCCACCATGCACTGGAA	CGCTTAA	264
QY	397	AAACGGCAATTGTAGT	413	
DB	265	AACCTCGATATTACTGT	281	

RESULT 4
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.

APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58

Query Match 7.6%; Score 34.6; DB 1; Length 456;
Best Local Similarity 53.3%; Pred. No. 0.098;
Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 277 ATGATAGCGGCGGAGTCCACGATGCAATTTGCAAGTGTCTTATGTAATCT 336
Db 208 ATTACCGAGCGGTTATGTAACGCGCGCATGTAGCGCGGATGCAATATAGTACT 267

Qy 337 GCGATGATTAATCCAGAAAGTTCTGTATTAAGCAATATTAACAGTATGTAATCA 396
Db 268 ATTAACGACTCTCAATGAGTTTCAGAAATATATGCAACATGCAACAGTGAACGCTAAA 327

Qy 397 AAAAGCGCAATTTGAGT 413
Db 328 AACTCGATATTACTGT 344

RESULT 5
US-08-545-528D-1
Sequence 1, Application US/08545528D
Patent No. 6537773
GENERAL INFORMATION:
APPLICANT: Frazer et al.
TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
PATENT NO. 6537773
FILE REFERENCE: P8193P1
CURRENT APPLICATION NUMBER: US/08/545,528D
CURRENT FILING DATE: 1995-10-19
PRIOR APPLICATION NUMBER: US 08/488,018
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/473,545

PRIOR FILING DATE: 1995-06-07
NUMBER OF SEQ ID NOS: 1
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 580073
TYPE: DNA
ORGANISM: Mycoplasma genitalium
US-08-545-528D-1

Query Match 7.0%; Score 31.8; DB 4; Length 580073;
Best Local Similarity 43.1%; Pred. No. 30;
Matches 153; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

Qy 26 TGTAAACAATCTGCGGCGCTGGAGTGAACCCGACGAGTATGATTTAGTAATT 85
Db 181203 TTTTAAATGTTTGGCACTAACAGTTAATTCAGTTTAAACAAAGAACAGTTACAGATG 181262

Qy 86 CAGAAATTAATCTGCGGCTAAATGAATTAAGTAAGTCTTCATTTATCAGGACGACATTA 145
Db 181263 CATGTGTAATCTTGGGTTGACCTTTAAGATGAAGAACTGAAGTTACCATTAATAATTTG 181322

Qy 146 TTGGTCAAGCTGGAGCTAATTAATAGTCTCAGTTACGCGAGGAGCTCAAACTTTTG 205
Db 181323 TAGCTAATCATCACTTTGAAGATAGTATGATTTATGCAACCAAGCACTTAATTTGTTA 181382

Qy 206 CGGTGTTGGCCAGAAAGTAGTAGCAACCGGCGAAAGATTGACACAGAGATTTATA 265
Db 181383 CTATCATGGGTGATGATGACCATGTAAACTTGGCTTTTAAACAAATTAAGAAACTTA 181442

Qy 266 ACCTTGATATATATGATCAGCGGCGAGTGCACAGATGCAATTTTGGCAAGTGCTT 325
Db 181443 ATGTAATCTGTAAGAGTTTGGGGAATTTACCCAAAATTTGGTCTTATCAGGTGAAA 181502

Qy 326 ATGTAATCTGCGCATATTAATTCAGAAAGTTCTGTAATTAAGCAATATTAC 380
Db 181503 ATCACAATAATTAACGATTACTTTATGTACTCTCGGCGATGAAGCATTTAC 181557

RESULT 6
US-09-790-988-1
Sequence 1, Application US/09790988
Patent No. 6632935
GENERAL INFORMATION:
APPLICANT: SHIGENOBU, SHUJI
APPLICANT: WATANABE, HIDEMI
APPLICANT: HATTORI, MASAHIRA
APPLICANT: SAKAKI, YOSHIYUKI
TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
FILE REFERENCE: 081356/0159
CURRENT APPLICATION NUMBER: US/09/790,988
CURRENT FILING DATE: 2001-02-23
PRIOR APPLICATION NUMBER: JP2000-107160
PRIOR FILING DATE: 2000-04-07
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 640681
TYPE: DNA
ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 7.0%; Score 31.8; DB 4; Length 640681;
Best Local Similarity 46.9%; Pred. No. 32;
Matches 99; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 2 TGAATAACAATGTTATTTATGATGTTAACAATAGTGGTGGCTGGATTCAGCCG 61
Db 79853 TTAGCAAGAAAGCTCGAGATTAATTTACACAAAAAGCATATGAACAGTTTC 79912

Qy 62 CAGCAGGTTATGATTTAGCTAATTCAGATTAATTAATTAATTAATTAATTAATTAATTA 121
Db 79913 AAACAGACTCTTAAGGCAAGAGAGAACTTGTGTAATAATTAATTAATTAATTAATTA 79972

QY 122 CTTCATTTATACGACGCCCATATTTGTCGACCTGGAGCTAATATATAGTCTCAAGTTAC 181
Db 79973 TATTTTAAATTTTGAATAATTCCTCTTGCTAAATTTGAATAATCAATATGTTCTCAATTAC 80032
QY 182 GGCACGAGCGCTCAAAACTTTGGCGGTGT 212
Db 80033 TAAACAAGCTTTTAAATTTATCTTATGT 80063

RESULT 7

```

US-09-107-532A-2325
/ Sequence 2325, Application US/09107532A
/ Patent No. 6583275
/
/ GENERAL INFORMATION:
/ APPLICANT: Lynn A Doucette-Stamm and David Bush
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
/ ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
/
/ NUMBER OF SEQUENCES: 7310
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: GENOME THERAPEUTICS CORPORATION
/ STREET: 100 Beaver Street
/ CITY: Waltham
/ STATE: Massachusetts
/ COUNTRY: USA
/ ZIP: 02354
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: CD-ROM ISO9660
/ COMPUTER: PC
/ OPERATING SYSTEM: <Unknown>
/ SOFTWARE: ASCII
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/107,532A
/ FILING DATE: 30-Jun-1998
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/085,598
/ FILING DATE: 14 May 1998
/ APPLICATION NUMBER: 60/051571
/ FILING DATE: July 2, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Ariniello, Pamela Deneke
/ REGISTRATION NUMBER: 40,489
/ REFERENCE/DOCKET NUMBER: GTC-012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (781)893-5007
/ TELEFAX: (781)893-8277
/
/ INFORMATION FOR SEQ ID NO: 2325:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 525 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: circular
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Enterococcus faecium
/
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (b) LOCATION 1...525
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2325:
US-09-107-532A-2325

```

```

RESULT 8
US-08-956-171E-133/c
/ Sequence 133, Application US/08956171E
/ Patent No. 6593114
/ GENERAL INFORMATION:
/ APPLICANT: Charles Kunach
/           Gil H. Choi
/           Patrick S. Dillon
/           Craig A. Rosen
/           Steven C. Barash
/           Michael R. Fannon
/ TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
/ NUMBER OF SEQUENCES: 5256
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Human Genome Sciences, Inc.
/ STREET: 9410 Key West Avenue
/ CITY: Rockville
/ STATE: Maryland
/ COUNTRY: USA
/ ZIP: 20850
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
/ COMPUTER: HP Vectra 486/33
/ OPERATING SYSTEM: MSDOS version 6.2
/ SOFTWARE: ASCII text
/
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/956,171E
/ FILING DATE: 20-Oct-1997
/ CLASSIFICATION: <Unknown>
/
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/009,861
/ FILING DATE: January 5, 1996
/ APPLICATION NUMBER: 08/781,966
/ FILING DATE: January 3, 1997
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Mark J. Hyman
/ REGISTRATION NUMBER: 46,789
/ REFERENCE/DOCKET NUMBER: PB248P1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (240) 314-1224
/ TELEFAX: (301) 309-8439
/
/ INFORMATION FOR SEQ ID NO: 133:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2785 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/
/ SEQUENCE DESCRIPTION: SEQ ID NO: 133:
US-08-956-171E-133

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2785;
Matches 80; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

125 CATTAAATCGGAGCGCATATTTGGTCAAGCTGAGGAGCTAATAATAGTGCTCAGTTAAGCGC 184
DB 2531 CAATTATCAAGACAGATATTCATTTCGAAACCGGTGCTCTTGTAAGTCTTAAGGTACCA 2472
QY 185 AGGAGAGCTCAAAACTTTTGGCGGTGTTGCGCAGAAAGGTAGTACCAACCGGGCAAGA 244
DB 2471 ATGATGTGCAACCTATTTATCTGCTTAATGATGATTAACCAAGCATAGGTGTTATACCAAAA 2412
QY 245 TTGACCAAGACAGAGATTTAATCCTTGATGATATATGATCAG 285
DB 2411 TTGCAACTGTATGCAATTGTATCTGCCAAAGTTAGACAG 2371

RESULT 9
US-09-801-861-3
/ Sequence 3, Application US/09801861
/ Patent No. 6482114

```


GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEROPOF
FILE REFERENCE: CL001098
CURRENT APPLICATION NUMBER: US/09/801,861
CURRENT FILING DATE: 2001-03-09
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 53332
TYPE: DNA
ORGANISM: Human
US-09-801-861-3

Query Match 6.9%; Score 31.4; DB 4; Length 53332;
Best Local Similarity 61.7%; Pred. No. 12;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 336 TGCATGATTTATCCGAAAGGTTCTGTATTAAGCAATTTACACGATGCTACTCA 395
Db 8632 TGTATGATGATTCACAAAGATATGTATTAAGCACTTATCACAGATCAGTTTTC 8691
Qy 396 AAAACGCAATTTAGTCA 416
Db 8692 GAGTTCAACAAATGTTGACCA 8712

RESULT 10
US-09-596-002-32/c
Sequence 32, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Paterson, Chandra
APPLICANT: Berg, Kim, L.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
FILE REFERENCE: PM-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
CURRENT FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 32
LENGTH: 62909
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: incyte template ID No. 6632636 32
PUBLICATION INFORMATION:
US-09-596-002-32

Query Match 6.9%; Score 31.4; DB 4; Length 62909;
Best Local Similarity 49.7%; Pred. No. 14;
Matches 80; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 221 AAGTAGTACGACCGGCAAGATTGACGACGAGATTAATTAACCTTCATATATTG 280
Db 8029 AATTGATTAACACCATATGAGATATGTATGAAGAAAGTTATCTTATTCATTTATG 7970
Qy 281 ATCAGCGCGGAGTGCACAGATGCAAGTATTTGCAAGTGTCTTATGTAATCTGCGA 340
Db 7869 CGGTGGCGAGTTTGTCTAACCGAGACTTATGAGCTTATATCATCACCAATGGC 7910
Qy 341 TGATTATCCAGAAAGTTCTGTATTAAGCAATATTACA 381
Db 7909 GTTTAGCATTTGACATTTTGGACCTTCACCAATGTCGA 7869

RESULT 11

US-09-540-236-1518
Sequence 1518, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
FILE REFERENCE: 2709, 2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 1518
LENGTH: 606
TYPE: DNA
ORGANISM: M. catarrhalis
US-09-540-236-1518

Query Match 6.8%; Score 31.2; DB 4; Length 606;
Best Local Similarity 50.0%; Pred. No. 1.5;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 226 AGTAGCAACCGGCAAGATTGACGACGAGATTAATTAACCTTCATATATTGATCAG 285
Db 4 ATTAACAACCATATGAGATATGTATGAAGAAATATCTTATTAATGATTTGCGCGTG 63
Qy 286 GCGGCGAGTGCACACGATGCGCATTTTCGCAAGTCTTATGTAATCTGCGATATT 345
Db 64 GCAGGTTTGTCTAACGACGACTTATGAGCTTATCTTATATCAACCAATGCCGTTT 123
Qy 346 ATCCAGAAAGTTCTGTATTAAGCAATATTACA 381
Db 124 AGCATTAACATTTTGGCAGCTCTCAACCAATGTGCA 159

RESULT 12
US-09-134-001C-398
Sequence 398, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 398
LENGTH: 1404
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-398

Query Match 6.8%; Score 31.2; DB 4; Length 1404;
Best Local Similarity 52.3%; Pred. No. 2.3;
Matches 69; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 275 ATATTGATCAGGCGGAGGCCAAGATGCGATTTTGGCAAGTGTCTTATGTAATA 334
Db 1079 ATTCTGATTAAGACGACGATGATGACTGTATTAATAGTCTTCAATCTTACCTACTT 1138
Qy 335 CTGCGATGATTTATCCAGAAAGTTCTGTATTAAGCAATATTACAAGATGTAATC 394
Db 1139 CAATTAATTAATACGATTAATGATGCAATATACTCAATTAATACCAATATCAC 1198
Qy 395 AAAAAGCGCA 406
Db 1199 AAGATTAATGCA 1210

```

RESULT 13
US-09-584-50
Sequence 50. Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Hartness, Robin B
APPLICANT: Loomore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESSES:
ADDRESS: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 2139 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-584-50

Query Match
Best Local Similarity 6.8%; Score 31; DB 4; Length 2139;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 306 CAGTATTCGACAGTGTGTTATGTTATGTCGATGATTCACGAAGTTCTGTTAA 365
DB 1845 CAGTATTCGACAGTGTGTTATGTTATGTCGATGATTCACGAAGTTCTGTTAA 365
QY 366 TAAAGCAATATTTACACAGTATGTATC 392
DB 1905 TAAAGCAATATTTACACAGTATGTATTAAC 1931

RESULT 14
US-08-916-421B-1/c
Sequence 1. Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-
patent No. 6503729
FILE REFERENCE: jannaschii
CURRENT APPLICATION NUMBER: US/08/916,421B

```

```

CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc.feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc.feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g

```

```

: LOCATION: (1349491)..(1349491)
: OTHER INFORMATION: n equals a, t, c, or g
: NAME/KEY: misc feature
: LOCATION: (1470091)..(1470091)
: OTHER INFORMATION: n equals a, t, c, or g
: NAME/KEY: misc feature
: LOCATION: (1569020)..(1569020)
: OTHER INFORMATION: n equals a, t, c, or g
: NAME/KEY: misc feature
: LOCATION: (1602912)..(1602912)
: OTHER INFORMATION: n equals a, t, c, or g
: NAME/KEY: misc feature
: LOCATION: (1603734)..(1603734)
: OTHER INFORMATION: n equals a, t, c, or g
: NAME/KEY: misc feature
: LOCATION: (1637998)..(1637998)
: OTHER INFORMATION: n equals a, t, c, or g
: NAME/KEY: misc feature
: LOCATION: (1664854)..(1664855)
: OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match
Best Local Similarity 6.8%; Score 31; DB 4; Length 1664976;
Matches 82; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 239 CAAGAATTGACACAGACAGATTAACTTGATATGATCAGCGCGGACGTCCA 298
DB 122859 CATTATATAGAGAGATGATGAACTACCGTGCTACTTATGACATGACAGAGACTG 122800
QY 299 ACGATGCCAGTATTTGCGCAGGTGCTTATGTTAATCTGCAGTATTATCCAGAAAGTT 358
DB 122799 GAGATTACATATTATGCAATAATATTGATGCTTACGCTGTTCAAAATAGCGATAAAT 122740
QY 359 CTGCTAATTAACCAATATTATACACGTTATGCTACTCAAAAAACGCCA 405
DB 122739 ATACAGTGGAGCATTAAACAGACGATGAGGCTTTTAGGAATGCA 122693

RESULT 15
US-09-621-976-2813
: Sequence 2813, Application US/09621976
: Patent No. 6639063
: GENERAL INFORMATION:
: APPLICANT: Dumas Milne Edwards, J.B.
: APPLICANT: Jobert, S.
: APPLICANT: Giordano, J.Y.
: TITLE OF INVENTION: ESTs and Encoded Human Proteins.
: FILE REFERENCE: GENSET.054PR2
: CURRENT APPLICATION NUMBER: US/09/621.976
: CURRENT FILING DATE: 2000-07-21
: NUMBER OF SEQ ID NOS: 19335
: SOFTWARE: Patent.pm
: SEQ ID NO 2813
: LENGTH: 832
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 235..399
US-09-621-976-2813

Query Match
Best Local Similarity 6.7%; Score 30.6; DB 4; Length 832;
Matches 69; Conservative 151; Mismatches 195; Indels 2; Gaps 1;

QY 10 AATATGTTATTTATGATGTTAAACATACACTGCGCGCTGGGATTCAGCGCAGCAGT 69
DB 40 WRRKKKAAWKKWTKMTWMMYRVAMWGYYKKKACRPTTKKKKKGGYMMWMMWGMRSYM 99
QY 70 TATGATTTAGCTTAATTCAGATATTAACCTTCCGGTAAATGAATAGTAAGTCTTCA 129
DB 100 AATATTTGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 159

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaataaacatgttctatc.....ctgcgtgacacacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243257 seqs, 1840798884 residues
Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	36	7.9	413	14 US-10-184-644-156 Sequence 156, App
2	36	7.9	413	14 US-10-184-634-156 Sequence 156, App
3	34.4	7.5	5184	14 US-10-247-671-55 Sequence 55, App1
4	33.8	7.4	1093	12 US-10-424-599-43422 Sequence 43422, A
5	33.8	7.4	1149	9 US-09-912-020-139 Sequence 139, App
6	33.8	7.4	1149	12 US-10-282-122A-6441 Sequence 6441, App
7	33.6	7.3	1299	12 US-10-282-122A-35109 Sequence 35109, A
8	33.2	7.3	1371	12 US-10-282-122A-17259 Sequence 17259, A
9	33	7.2	312	12 US-10-424-599-104721 Sequence 104721, A
10	33	7.2	1194	15 US-10-369-493-32803 Sequence 32803, A
11	33	7.2	1269	12 US-10-282-122A-35991 Sequence 35991, A
12	33	7.2	1837	12 US-10-282-122A-35601 Sequence 35601, A
13	33	7.2	319630	15 US-10-398-221-7 Sequence 7, App1
14	33	7.2	3011208	15 US-10-398-221-2058 Sequence 2058, App
15	32.8	7.2	746	15 US-10-027-632-14807 Sequence 14807, A

c 16	32.8	7.2	746	15	US-10-027-632-14808	Sequence 14808, A
c 17	32.8	7.2	2552	9	US-10-027-632-111939	Sequence 111939, A
c 18	32.6	7.1	351	9	US-09-983-965-4342	Sequence 4342, App
c 19	32.6	7.1	2000	15	US-10-260-238-1758	Sequence 1758, App
c 20	32.6	7.1	2250	15	US-10-369-493-46046	Sequence 46046, A
c 21	32.6	7.1	17419	14	US-10-339-676-100	Sequence 100, App
c 22	32.6	7.1	17419	14	US-10-311-455-1268	Sequence 1268, App
c 23	32.6	7.1	17419	14	US-10-240-453-112	Sequence 112, App
c 24	32.4	7.1	1320	12	US-10-282-122A-10004	Sequence 10004, A
c 25	32.4	7.1	2379	12	US-10-282-122A-9449	Sequence 9449, App
c 26	32.2	7.0	2634	12	US-10-282-122A-35828	Sequence 35828, A
c 27	32	7.0	2565	15	US-10-369-493-25168	Sequence 25168, A
c 28	31.8	7.0	409	9	US-09-864-761-1955	Sequence 2955, App
c 29	31.8	7.0	438	9	US-09-864-761-19736	Sequence 19736, A
c 30	31.8	7.0	708	14	US-10-184-644-584	Sequence 584, App
c 31	31.8	7.0	708	14	US-10-184-634-584	Sequence 584, App
c 32	31.8	7.0	1860	10	US-10-282-122A-27280	Sequence 27280, A
c 33	31.8	7.0	11138	10	US-09-960-870-5	Sequence 5, App1
c 34	31.8	7.0	11138	10	US-09-960-858-5	Sequence 5, App1
c 35	31.8	7.0	580073	14	US-10-205-220-1	Sequence 1, App1
c 36	31.8	7.0	640681	9	US-09-790-988-1	Sequence 1, App1
c 37	31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, App
c 38	31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, App
c 39	31.6	6.9	575	15	US-10-027-632-222766	Sequence 222766, App
c 40	31.6	6.9	2565	15	US-10-321-802-19	Sequence 19, App1
c 41	31.4	6.9	867	15	US-10-027-632-167010	Sequence 167010, App
c 42	31.4	6.9	867	15	US-10-027-632-167011	Sequence 167011, App
c 43	31.4	6.9	867	15	US-10-027-632-167012	Sequence 167012, App
c 44	31.4	6.9	1407	12	US-10-424-599-25067	Sequence 25067, App
c 45	31.4	6.9	2785	8	US-08-781-986A-133	Sequence 133, App

ALIGNMENTS

RESULT 1

US-10-184-644-156

Sequence 156, Application US/10184644

Publication No. US200304930A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Goddard, Paul J.

APPLICANT: Gurney, Austin L.

APPLICANT: Pan, James

APPLICANT: Smith, Victoria

APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.

APPLICANT: Zhang, Zhenli

TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

FILE REFERENCE: P3430R1C227

CURRENT APPLICATION NUMBER: US/10/184, 644

CURRENT FILING DATE: 2002-06-28

Prior Application removed - See File Wrapper or Palm

NUMBER OF SEQ ID NOS: 612

SEQ ID NO 156

LENGTH: 413

TYPE: PRT

ORGANISM: Homo Sapien

US-10-184-644-156

Query Match

Best Local Similarity 11.6%; Pred. No. 2;

Matches 28; Conservative 93; Mismatches 121; Indels 0; Gaps 0;

147 TGGTCAAGCTGGACATATATAGTGTCACTTACGCGAGGAGCTCAAAATTTTGGC 206

172 HDKDYKSGSKRVRGLTKRRNSGSKRSGRSRRASGQDRBTREHLBRKAGGR 231

207 GGTTGTTGCCAAGAAGTAGTACCAACGGCGAAGATTGACCAACAGAGATTATTA 266

[illegible]

```

RESULT 2
US-10-184-634-156
; Sequence 156, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Yilan
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184, 634
; Prior Application removed - 2002-06-28
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 156
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-184-634-156

```

Query Match	7.9%	Score 36;	DB 14;	Length 413;
Best Local Similarity	11.6%	Pred. No. 2;		
Matches	28;	Conservative	93;	Mismatches 121; Indels 0; Gaps 0;
Qy	147	TGCTCAAGCTGGAGCATTAATTAATAGTCTCAGTTACGGACGGAGGCTCAAAACTTTGGC		
Db	172	HGCKDLYVKSCKLRYVLLMKRNKSGSKRRGSKRRRRASSGDQREGREHLDRRAQGR		
Qy	207	GGTGTGTGGCGAAGAAGGTAGTAGCAACCGGGCAAAAGTTACCAGACAGAGATTAA		
Db	232	RKKKSGRGRIAREGRPSFQWTRVYKNTIHPKMAAGMGEDATLTDYALBELKPAHKKKVM		
Qy	267	CCCTTCATATATTTGATTCAGCGGGGCGAGTCCCAACGATGCGAGTATTTCGCAAGTCTTA		
Db	292	ELGISPTIKKMPGCMHPSFSDNDRADOLVYRFGCSVDSDENDLLYQCAESESSTSGVY		
Qy	327	TGCTAATATCTCGATGATTATCCAGAAGGTCTCGGTAATATAAAGCAAAATTTACAACAGTA		
Db	352	LRLDPPKWKRRKIIVVSGHGVNDVHGQKDYNAVATVITPLKYAQLICLITGNDANCA		
Qy	387	TG 388		
Db	412	YG 413		

RESULT 3
US-10-247-671-55/c
' Sequence 55, Application US/10247671

```

1 Publication No. US20030194721A1
2
3 GENERAL INFORMATION:
4
5 APPLICANT: Mikita, Thomas
6
7 APPLICANT: Shiffman, Dov
8
9 APPLICANT: Porter, Gordon, J.
10
11 APPLICANT: Kaser, Matthew R.
12
13 TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
14
15 FILE REFERENCE: PA-0050 US
16
17 CURRENT APPLICATION NUMBER: US/10/247,671
18
19 CURRENT FILING DATE: 2002-09-18
20
21 PRIOR APPLICATION NUMBER: 60/323,784
22
23 PRIOR FILING DATE: 2001-09-19
24
25 NUMBER OF SEQ ID NOS: 186
26
27 SOFTWARE: PERL Program
28
29 SEQ ID NO 55
30
31 LENGTH: 5184
32
33 TYPE: DNA
34
35 ORGANISM: Homo sapiens
36
37 FEATURE:
38
39 NAME/KEY: misc feature
40
41 OTHER INFORMATION: Incyte ID No. US20030194721A1 5185743CB1
42
43 US-10-247-671-55

```

Query Match	Best Local Similarity	7.5%;	Score 34.4;	DB 14;	Length 5184;
Matches	74;	Conservative	0;	Mismatches 66;	Indels 0;
					Gaps 0
Qy	51	GATTGAGCGCGAGAGGTTATGATTTAGCTAATTCAGATATATACCTTCGCGTAAATCA	110		
Db	2489	GCCTTGAGCTGCTGTAATAGTATCTGGAAGCTACTACACTATTGTGATAGAAAAATA	2430		
Qy	111	ATTGAGTAAGTCTTCATTATATCAGGAGGACCAATATTTGCTCAAGCTGGAGCAATATATAG	170		
Db	2429	TTTGATGAAATCTGCTGCTGAAGTCTGTGCTTGTCTTAGTGGTGAGACGCTGCTGT	2370		
Qy	171	TGCTAGATTACGAGGAGAG	190		
Db	2369	TTTGTGATGCTTCAGTCAG	2350		

```

RESULT 4
US-10-424-599-43422/c
; Sequence 43422, Application US//10424599
; Publication No. US20040031072A1
GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223) B
CURRENT APPLICATION NUMBER: US/10/424, 599
NUMBER OF SEQ. ID NOS: 2003-04-28
SEQ ID NO 43422
; LENGTH: 1093
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_139209C.1
US-10-424-599-43422

```

Query Match	Best Local Similarity	Score	DB	Length
Matches 59; Conservative	7.4%; 58.4%;	33.8;	12;	1093;
		Pred. No. 15;		
		Mismatches 42;	Indels 0;	Gaps 0
QY	294	TCGCAACGATGCAGTATTTCGACAGTCTTATGTAATCTGCATATTATTCAGAA	353	
Db	640	TGCAACAGTTGAGATCATCAGAGGTGTAAGTTTCCGACAGATTTTGCAGAA	581	
QY	354	AGTCTTCGTATTAAGCAATATTACACAGTATGTACTC	394	

Db 580 AGATTCTGTGTAGTCTGCAGTTACAGCTAATGTAATGTC 540

RESULT 5

US-09-912-020-139

Sequence 139, Application US/09912020
Patent No. US20020045592A1

GENERAL INFORMATION:

APPLICANT: Zykkind, Judith
APPLICANT: Ohlsen, Karl L.
APPLICANT: Trawick, John
APPLICANT: Forsyth, R. Allyn
APPLICANT: Froelich, Jamie M.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
FILE REFERENCE: ELITRA.001DVI

CURRENT FILING DATE: 2001-07-23

PRIOR APPLICATION NUMBER: 09/492,709

PRIOR FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: 60/117,405

PRIOR FILING DATE: 1999-01-27

NUMBER OF SEQ ID NOS: 485

SOFTWARE: PastSeq for Windows Version 3.0

SEQ ID NO 139

LENGTH: 1149

TYPE: DNA

ORGANISM: E. Coli

US-09-912-020-139

Query Match 7.4%; Score 33.8; DB 9; Length 1149;
Best Local Similarity 56.9%; Pred. No. 16;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 259 GATTATACCTTGATATATGATGAGCGGCGGCAAGATGCGAATTTGCGAA 318

Db 442 GATATTAACCTGGAATTAATGAGTTCGCTACCGTACCCACGAGGTAATACATCGTCT 501

Qy 319 GGTGCTTATGCTAATCTGCGATGATTTATCCAGAAAGTTCTGTAAATA 367

Db 502 GGCACACAGTGGTTACAGTCTCTTAATATATGCTGAGCTTAATGTAATA 550

RESULT 6

US-10-282-122A-6441

Sequence 6441, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6441
LENGTH: 1149
TYPE: DNA
ORGANISM: Escherichia coli
US-10-282-122A-6441

Query Match 7.4%; Score 33.8; DB 12; Length 1149;
Best Local Similarity 56.9%; Pred. No. 16;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 259 GATTATACCTTGATATATGATGAGCGGCGGCAAGATGCGAATTTGCGAA 318

Db 442 GATATTAACCTGGAATTAATGAGTTCGCTACCGTACCCACGAGGTAATACATCGTCT 501

Qy 319 GGTGCTTATGCTAATCTGCGATGATTTATCCAGAAAGTTCTGTAAATA 367

Db 502 GGCACACAGTGGTTACAGTCTCTTAATATATGCTGAGCTTAATGTAATA 550

RESULT 7

US-10-282-122A-36109

Sequence 36109, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

[illegible]

		Matches	Indels	Gaps
Oy	82	AATTCAGATATTAACCTTGGCGGTAAAGAAATTGATGACTCTTCATTATACAGGACGCC	103	0
Db	466	AATTAAGAAACATGAGTGGCTGTATATGAGATGGGTATGTGCAATTCATAGAAATPAAA		
Oy	142	ATATATGTCACACCTGGGACTATATATATATAGTCTCAGTTACGGCAGGAGGCTCAAACTT		
Db	526	TATTTAGTGGATATAGTAAATCCTPAAATTTGGACGTCATATGAAATATAGATTAATCAGT		
Oy	202	TTCGCGCTGTTGGCAGAAGGTAATACCAACCGGGCAAAAGTTGACCGACAGAAAT		
Db	586	GTGAGAAAATACAGGTTCCACAGAGGATATTTTGAAGCAAGATGAGATATCTTCAAAC		
Oy	262	TATTAACCTTGATATAT	279	
Db	646	TTTGATGAACAAATACT	663	

APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 32803
LENGTH: 1194
TYPE: DNA
ORGANISM: Chloroflexus aurantiacus
US-10-369-493-32803

Query March 7.2% Score 33; DB 15; Length 1194;
Best Local Similarity 48.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 167 ATAGTCTCAGTTACGGGAGGAGGCTCAAACTTTGGCGGTTGTCGCAAGAAGTA 226
DB 737 ACACGTCTCGGTTGCACGAGGAAATCACCAACATTCAGCCCTTGGCAAGCAAGAGGTG 678
QY 227 GTACCAACCGGCAAGATTGACACAGAGAGATTATACCTTGACATATATTGATCAG 286
DB 677 TTGGCCATACCAACCAATGAGGATGATGATGATGATGATGATGATGATGATGATG 618
QY 287 CGGGACGCGCAAGATGCTGATTTTGGCAAGGCTTATGATGATGATGATGATGATGATG 346
DB 617 CCGATCTTGCTGATGATTTTGGCAACGATGAGGATGAGGATGAGGATGAGGATGAGGATG 558
QY 347 TCCAG 351
DB 557 TTCAG 553

RESULT 11

US-10-282-122A-35991
Sequence 35991, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35991
LENGTH: 1269
TYPE: DNA
ORGANISM: Streptococcus mutans
US-10-282-122A-35991

Query March 7.2% Score 33; DB 12; Length 1269;
Best Local Similarity 49.2%; Pred. No. 29;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 236 GGGCAAGATTGACACAGAGAGATTATTAACCTTGATATATGATCAGCGGCACTG 295
DB 413 GAGCAAGATTCAACAGACAGGTGTGATATTAACGACCTGCCATGCTTTAAAGGTG 472
QY 296 CCAAGATGCCGATTTTGGCAAGGCTTATGATTAATCTGCATGATTAATCAAGAA 355
DB 473 CTAACATTTACATGATTTTCCAGTGTGGCGGACTCAGAACTTATGATGCTGCTA 532
QY 356 GTTCTGTAATAAGCAATATTACAGATGATGATCTCAAAAAACGCAATTTAG 412
DB 533 CTTAGCTGATGAGAACGATTAATGCAAAATGACAGCTGAACCAAGAAATCGTTG 589

RESULT 12

US-10-282-122A-35601
Sequence 35601, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614

Best Local Similarity 50.6%; Pred. No. 25;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```
OY      239 CAAAGATTGACCAAGACAGAGATTATTAACCTTGCAATATTTGATCAGCGGGCAAGTCCA 298
      |||||
Db      746 CAGAGACTTTCCTCCACTGTGTATTAATAAGATTCCATGTGATAGCTTGCTGGCAGCCCC 687
      |||||
OY      299 ACGATGCCAGTATTTGCGAAGTGCTTATGGTAATACTGCGATGATTATCCAGAAAGTT 358
      |||||
Db      686 ATGAGGCCGTGATGACTCTTAAGACGACTTGTAGAACTGCCATTTCAATGGCTAGAGTTT 627
      |||||
OY      359 CTGGTAATPAAGCAAAATATTACACAGTATGTAATC 394
      |||||
Db      626 TAATTAATTTTGGATCTTTGAAAAATAGTGAATC 591
      |||||
```

Search completed: March 17, 2004, 08:15:26
Job time : 408.736 secs

THIS PAGE BLANK (USPTO)

Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
DNA Sequencing by: Washington Song
Source lab clone id - 374663
Trace considered overall poor
royalty-free through NHLBI. Contact the IMAGE Consortium
(image@image.llnl.gov) for further information.
Seq primer: -40RP from Gibco
High quality sequence stop: 1

Bouyce

1. 220
/organism="Xenopus laevis"
/db_type="mRNA"
/db_xref="taxon:3815"
/title="IMAGE:3749963"
/tissue_type="gastrocn
/lab_host="Top-10 P"
/clone_id="Xenopus laevis gastrula non normalized"
XhoI; DNA was prepared from 2 ug of poly A+ RNA (equal
parts from stage 10.5 and stage 11.5 gastrulae).
EcoRI-XhoI cut DNA was then ligated into UniZap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
Top10P. Clones were picked excised and used to infect
litter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM
KH2PO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4
7H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 2
hours. Original library construction by Bruce Blumberg
(Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Smith College Department of Biological Sciences
Department of Biological Sciences, Clark Science Center, Smith
College, Northampton, MA, 01063, USA
Tel: 4135853826
Fax: 4135853786
Email: genome@smith.edu
Seq primer: Bluescript SK.
Location/Qualifiers
1. .207
/organism="Onchocerca volvulus"
/mol_type="mRNA"
/db_xref="taxon:6282"
/clone="SMOVMFCAR04G09"
/dev_stage="microfilaria"
/lab_host="X1-Blue MRF"
/clone_1b="Onchocerca volvulus microfilaria cDNA
(SMW98MTL-OWME)"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
prepared from approximately 200,000 microfilariae isolated
from the skin of infected individuals from Kinshasa,
Cameroon and converted to double-strands from Kinshasa,
reverse transcriptase and oligo(dT) followed by RNase H
and DNA pol I. The library has 7.8 x 10⁵ independent
1kb. The library was constructed by Michelle
Lizotte-Waisviski. The library is available from
Dr. S.A. Williams, email:genome@smith.edu."

FEATURES
Source

Location/Qualifiers
1. .207

Query Match	45.1%;	Score 205.8;	DB 12;	Length 220;
Best Local Similarity	96.8%;	pred. No. 4.7e-51;		
Matches 210; Conservative				

QY 80 CTAATTCAGATATATACCTCGCGGTAAGAATTAGTAGTCTCTTCAATTATACAGCAG 139
 Db 3 CTAATTCATATATATACCTTCGCGGCACATGATCAGTAAGTCTCTTCAATTATACAGCAG 139
 QY 140 CCATATATTGGTCACAGCTGGGACTAAATATATAGTGTCAAGTTACGCAAGGAGGCTCAAAA 199
 Db 63 CCATATATTGGTCACAGCTGGGACTAAATATATAGTGTCAAGTTACTGAGGGAGGCTCAAAA 199
 QY 200 TTTTGGGGGCTGTGTGGCAGAAGGTAGTACCAACGGGGCAAAAGATTGACCGACAGAG 122
 Db 123 TTTTGGGGGCTGTGTGGCAGAAGGTAGTACCAACGGGGCAAAAGATTGACCGACAGAG 259
 QY 260 ATTATACCTTGATATATATTGATCAGCGCGGACAGTGC 296
 Db 183 ATTATACCTTGATATATATTGATCAGCGCGGACAGTGC 219

Query Match
Best Total

Best Local Similarity	12.5%;	Score 56.8;	DB 9;	Length 207,
Matches 58;	Pred. 95.1%;	Pred. No. 3.9e-06;		
Conservative	0;	Mismatches 3,		

		C ₁	C ₂	indels	Gaps	
Oy	1	ATGAAAAACAAATTGTATTATGATGTTAAACAATACTGGCGTCCGCTGGAGATTGCACC			0	
Db	61	ATGAAAACCAATTCATTCATTATGATGTTAACAACTACTGGTGCCCTGGAGATTGCACC			60	
Oy	61 G	61				
Db	1 G	1				

RESULT 3

Bu618668/c	Bu618668	680 bp	mRNA	linear	EST 01-OCT-2003
LOCUS					
DEFINITION	Bu618668	NIBB Mochi	normalized	Xenopus early gastrula	library
ACCESSION	Bu618668	Xenopus laevis	CDNA clone	Xl186b22 5',	mRNA sequence.
VERSION	Bu618668				

ORGANISM

Xenopus laevis (African clawed frog)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Amurota, C. Tanaka, Y. Verrebrats, Euteloostomi
Xenopodinae, Xenopus.
1 (baes 1 to 680)
Kitsayama, A., Teratsaka, C., Mochii, M., Ueno, N., Shun-I, T. and
Kohara, Y.
Expressed genes in *X. laevis* embryo
Unpublished (2001)
Contact: Tadatsu Shin

COMMENT

Contact: Tadjan Shiri

Yata, Mishima, Shizuoka 411-8540, Japan

the information of this clone

FEATURES
SOURCE

Location/Qualifiers
1. 680
http://xenopus.nibb.ac.jp.

```

/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186D2"
/tissue_type="whole embryo"
/dev stage="stage 10.5"
/clone_lib="NIBB Mochi normalized Xenopus early gastrula
library"

```

ORIGIN

```

Query Match      10.3%; Score 47; DB 12; Length 680;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 410 TAGTGCAGACGACGCGCAATGCTATTCGGCTGACACACGTTAA 456
      |||||
Db 680 TAGTGCAGACGACGCGCAATGCTATTCGGCTGACACACGTTAA 634

```

RESULT 4

```

LOCUS      CNS04QW8      956 bp      DNA      linear      GSS 01-SEP-2000
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ORI end of clone
              129D18 of library G from Tetradon nigroviridis, genomic survey
              sequence.
ACCESSION  AL302777      GI:8183119
KEYWORDS   GSS; genome survey sequence.
SOURCE     Tetradon nigroviridis
ORGANISM   Tetradon nigroviridis

```

KEYWORDS

```

Tetradon nigroviridis
Tetradon nigroviridis
Tetradon nigroviridis

```

SOURCE

```

Tetradon nigroviridis
Tetradon nigroviridis
Tetradon nigroviridis

```

ORGANISM

```

Tetradon nigroviridis
Tetradon nigroviridis
Tetradon nigroviridis

```

REFERENCE

```

1 Roest Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
  Bernot,A., Filames,C., Winkler,P., Brotlier,P., Quetier,F.,
  Saurin,M. and Weissenbach,J.
  Estimate of human gene number provided by genome-wide analysis
  using Tetradon nigroviridis DNA sequence
  Nat. Genet. 25 (2), 235-238 (2000)

```

TITLE

```

JOURNAL MEDLINE 20296633
PUBMED 10835645

```

REFERENCE

```

2 Roest Crolius,H., Jallion,O., Dasilva,C., Ozouf-Costaz,C.,
  Filames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
  Saurin,M., Bernot,A. and Weissenbach,J.
  Characterization and repeat analysis of the compact genome of the
  freshwater pufferfish Tetradon nigroviridis
  Genome Res. 10 (7), 939-949 (2000)

```

TITLE

```

JOURNAL MEDLINE 20359837
PUBMED 10899143

```

REFERENCE

```

3 (bases 1 to 956)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqef@genoscope.cns.fr
http://www.genoscope.cns.fr/Tetradon.
Location/Qualifiers
1. 956
/organism="Tetradon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="129D18"
/clone_lib="G"
/notes="Genoscope sequence ID : COBGI29D809SP1-end :
PUC-ORI"

```

COMMENT

```

This sequence is a single read and was generated as part of a large
scale clone and sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.

```

FEATURES

```

source

```

ORIGIN

```

Query Match      8.8%; Score 40.2; DB 29; Length 956;
Best Local Similarity 46.6%; Pred. No. 0.62;
Matches 129; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

```

```

Cy 113 TGAGTAAGCTCTTCAATTTATCAGGACGACCAATATTTGTCAGACTGGGACTATATAGTG 172
      |||||
Db 430 TCAGTACGTGAGGAGCCACTATATCATCATGATCAGTCAGTCAGGGGTCTACTATCAT 371
      |||||
Cy 173 CTCAGTTACGCGAGGAGGCTCAAACTTTGGCGGTGTGGCGAAGGTACTAGCA 232
      |||||
Db 370 CACTGATCAGTCAGTCAGGAGCCACTATATCATCATGATCAGTCAGTCAGTCAGTCA 311
      |||||
Cy 233 ACCGGGCAAGATTGACGACAGGAGTTATTAACCTTGATATATATGATCAGGGGCA 292
      |||||
Db 310 ATCAATCAGTCATGATGATGAGAGGAGTCACTATATCATCTGATCAGTCAGTACGAG 251
      |||||
Cy 293 GTGCAACGATGCGAGTATTTCCGAAAGTGTATGATATATGATGATATATCAGA 352
      |||||
Db 250 TCACATATCATCATCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 191
      |||||
Cy 353 AAGCTTCTGTTAATTAAGCAATATTTACACATATG 389
      |||||
Db 190 CAGGAGCCACTATATCATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 154
      |||||

```

RESULT 5

```

LOCUS      BB703456      533 bp      mRNA      linear      EST 11-OCT-2001
DEFINITION BB703456 RIKEN full-length enriched, in vitro fertilized eggs Mus
              musculus CDNA clone 7420452F19 3', mRNA sequence.
ACCESSION  BB703456      GI:16052291
KEYWORDS   EST.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus

```

REFERENCE

```

1 (bases 1 to 533)
Mammalia: Eutheria; Chordata; Vertebrata; Euteleostomi;
Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

TITLE

```

JOURNAL MEDLINE 20359837
PUBMED 10899143

```

REFERENCE

```

2 Roest Crolius,H., Jallion,O., Dasilva,C., Ozouf-Costaz,C.,
  Filames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
  Saurin,M., Bernot,A. and Weissenbach,J.
  Characterization and repeat analysis of the compact genome of the
  freshwater pufferfish Tetradon nigroviridis
  Genome Res. 10 (7), 939-949 (2000)

```

TITLE

```

JOURNAL MEDLINE 20359837
PUBMED 10899143

```

REFERENCE

```

3 (bases 1 to 956)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 Evry cedex - FRANCE (E-mail : seqef@genoscope.cns.fr
http://www.genoscope.cns.fr/Tetradon.
Location/Qualifiers
1. 956
/organism="Tetradon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="129D18"
/clone_lib="G"
/notes="Genoscope sequence ID : COBGI29D809SP1-end :
PUC-ORI"

```

COMMENT

```

This sequence is a single read and was generated as part of a large
scale clone and sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.

```

FEATURES

```

source

```

ORIGIN

```

PUC-ORI"

```


REFERENCE Trichinelidae; Trichinella.
1 (bases 1 to 769)
AUTHORS Daub J., Connolly, B., Garate, T. and Blaxter, M.L.
TITLE A survey of genes expressed in the muscle stage larvae of the
parasitic nematode Trichinella spiralis
JOURNAL Unpublished (2001)
COMMENT Contact: Blaxter M.
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
PCR Primers
FORWARD: T7SEQ (GTAATACGACTCTATAGG)
BACKWARD: M13 Forward (CGCCAGGGTTTCCAGTCACGAC)
Seq primer: T7SEQ (GTAATACGACTCTATAGG).

FEATURES
source
1..769
Location/Qualifiers
/organism="Trichinella spiralis"
/mol_type="mRNA"
/db_xref="taxon:6334"
/clone="MBTSM1A133"
/sex="mixed"
/dev_stage="muscle stage larvae"
/note="Trichinella spiralis muscle stage larvae (BC)"
Site 2: Not (3'end); The infective L1 larva of
Trichinella spiralis is a nematode parasite of mammalian
skeletal muscle. The library was constructed using muscle
larvae of the T. spiralis isolate 1553, and was provided
by Dr Bernadette Connolly, University of Aberdeen."

ORIGIN

Query Match
Best Local Similarity 8.0%; Score 36.4; DB 12; Length 769;
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 239 CAAGATTGACGACGAGATTATTAACCTTGCATATTTGATGATGAGCGGCGAGTCCCA 238
DB 78 CAGTTATGAATTCGATTAACGAAATATATAGCGCTTCAAAATTGGTCATCGACTTCA 237
QY 299 ACGATGCCAGTATTTGCAAGGCTTATGATGATTAATCTGCATGATTAATCCAGAAAGCTT 358
DB 138 TCATGCAAGCATTTGATGAAGATTTTACCTATGACTGATTAATGATTAAGCTTTGTTT 197
QY 359 CTGATATTAAGCAATATTACAGATGATGATCAAAAACGCAATTGAG 412
DB 198 CTGATATTAAGCAATATTGTTATGAAGTGAAGAAATCAATGATTAATGAG 251

RESULT 14
LOCUS AZ811111 491 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0077A11F Mouse 10kb plasmid UGCGM library Mus musculus genomic
clone UGCG2M0077A11 F, genomic survey sequence.
ACCESSION AZ811111
VERSION AZ811111.1 GI:12979042
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 491)
AUTHORS Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Reilly, M., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T.,
Niederhausen, A., and Wright, D., Weis, R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weis

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: dunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0077 row: A column: 11
Seq primer: CGTTGTAACGACGCGCCACT
Class: plasmid ends
High quality sequence stop: 491.
Location/Qualifiers
1..491
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG2M0077A11"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PMD42nv; Purified genomic DNA from M.
laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
ligated to the blunt ends in high molar excess. The
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g14732114[9b]AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptor mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN

Query Match
Best Local Similarity 7.9%; Score 36.2; DB 28; Length 491;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 270 TCATATATTAATGACGGCGGCGGCAAGATGATTTGCAAGTGCTTATG 329
DB 110 TGAGATTAATATTTTCAGAGCTGTGTCAGCAAGGCGATATATCTTCAATGTTAAG 169
QY 330 TAATACGCCATGATTTTCCAGAAAGTCTGTAATAAGCAATATTACACGATG 389
DB 170 TACATGTCAGATGTTCTGTAATATGATGTTATCAAGAGAAAGAAAAAGAA 229
QY 390 TACTCAAAAACGCAATTTGTGTGACAGAGCA 422
DB 230 AAAAAAGAAAAAGAAAAAGAAATGAAAAAGAA 262

RESULT 15
LOCUS AZ811109 532 bp DNA linear GSS 20-FEB-2001
DEFINITION 2M0077A09F Mouse 10kb plasmid UGCGM library Mus musculus genomic
clone UGCG2M0077A09 F, genomic survey sequence.
ACCESSION AZ811109
VERSION AZ811109.1 GI:12979038
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 532)

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Rellly,M., Rose,R., Rose,R., Stokes,R., Tingey,A., von
Niedermauern,A. and Wright,D.,Weiss,R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT
84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0077 row: A column: 09

Seq primer: CTTGTAAACGACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 532.

FEATURES

Source

1..532
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUC2M0077A09"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/clone_lib="Mouse 10kb plasmid UUC1M library"
/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (g14732114[gb|AF129072.1]), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN

Query Match 7.9%; Score 36.2; DB 28; Length 532;

Best Local Similarity 52.3%; Pred. No. 8.4;

Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

QY 270 TGCATATATTGTCAGCGCGGAGTCCAGATCCGATATTCGCAAGTCTTATGG 329
    |||||
DB 92 TGAGGATTAATATTTTCAGAGCTGTGTCAAGCAAGGCGATATCGTTTCAATGTTAGAG 151
    |||||
QY 330 TAATACGTCGATGATTAATCCAGAAAGTTCTGTAATAAAGCAATATTACACAGTATGG 389
    |||||
DB 152 TACATGTCCGATAGTCTCGAAATATATGATTTGATACAGAAAGAAAGAAAGAAAGAA 211
    |||||
QY 390 TACTCAAAAAACGCAATTTGATGTCAGAGACA 422
    |||||
DB 212 AAAAAGAAAAAGAGAAAAAGATAGAAAAAGAAA 244
    |||||

```

Search completed: March 16, 2004, 04:28:52
Job time : 2237.91 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 206.439 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48

Sequence: 1 tatgatcagctggtaccg.....cccatgaatgagcatgca 48

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```
Database :
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_em:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_gst:*
12: gb_gst:*
13: gb_un:*
14: gb_vt:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vt:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pla:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_ey:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29.4	61.2	1797	3	LEIMSP52A
2	29.4	61.2	1860	3	LIGP63GEN
3	29.4	61.2	2104	3	LDO495002
4	29.4	61.2	2105	3	LDO495006
5	29.4	61.2	2106	3	LDO495007
6	29.4	61.2	2107	3	LDO495005
7	29.4	61.2	2109	3	LDO495003
8	29.4	61.2	2892	3	LEIGP63D
9	29.4	61.2	2966	3	LEIGP63B
10	29.4	61.2	3047	3	LEIGP63A
11	29.4	61.2	3047	3	LHU48798
12	29.4	61.2	3105	3	LEIGP6A
13	29.4	59.2	2621	3	LEIGP63C
14	28.4	57.9	2050	3	LTU495010
15	27.8	57.9	2058	3	LAES52831
16	27.8	57.9	2077	3	LTR495009
17	27.8	57.9	2090	3	LARS52830
18	27.8	57.9	2125	3	LTR495008
19	27.8	57.9	2161	3	LMGP63
20	27.8	57.9	3229	3	LMGP63C1
21	27.8	57.9	11523	1	AE014950
22	27.8	57.9	11523	1	AE014950
23	27.8	57.9	57112	2	AC019976
24	26.8	55.8	88031	2	AC004378
25	26.8	55.8	124700	3	AC005558
26	26.8	55.8	126807	9	HS391022
27	26.8	55.8	171774	3	AC009909
28	26.8	55.8	295134	3	AE003582
29	26.8	55.0	1800	3	LEIMSP54A
30	26.4	54.6	2046	3	LEIGP63E
31	26.2	53.8	238365	2	AC107269
32	25.8	51.3	1964	3	LIGP63
33	24.6	51.3	2544	6	AX568314
34	24.6	51.3	3636	6	AR218846
35	24.6	51.3	3636	6	BD003758
36	24.6	51.3	11495	1	AE007386
37	24.6	51.3	12092	1	AE008447
38	24.6	51.3	147006	2	SPNEU1904
39	24.6	51.3	349980	6	AX571761
40	24.6	51.3	349980	6	AX571762
41	24.6	50.8	2240	8	AY118173
42	24.4	50.4	144239	10	AL627347
43	24.2	50.4	182340	2	AC099639
44	24.2	50.4	223974	2	AC106343
45	24.2	50.4	223974	2	AC106343

ALIGNMENTS

RESULT 1
LOCUS LEIMSP52A 1797 bp DNA linear INV 28-NOV-1994
DEFINITION Leishmania donovani glycoprotein 63 (mgs2) gene, complete cds.
VERSION L19563.1 GI:308887
ACCESSION L19563
KEYWORDS glycoprotein 63, surface protein, virulence factor, zinc protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 1797)
AUTHORS Roberts,S.C., Swihart,K.G., Agey,M.W., Ramamorthy,R., Wilson,M.E.
and Donelson,J.B.
TITLE Sequence diversity and organization of the msp gene family encoding

JOURNAL M63 of Leishmania chagasi
MOL. Biochem. Parasitol. 62 (2), 157-171 (1993)
MEDLINE 94187792
PUBMED 8139613

COMMENT

Original source text: Leishmania donovani (sub_species chagasi)
stationary stage promastigote DNA.

FEATURES

Location/Qualifiers

1..1797
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
/dev_stage="stationary stage promastigote"
/gene="msp2"
1..1797
/gene="msp2"
1..1797
/note="propeptide"
/codon_start=1
/product="glycoprotein 63"
/protein_id="AA53688.1"
/db_xref="GI:308888"
/translation="MSVSSSTHRRSVARLVRLAAGAIVAAVGTAAAMAHAGV
OHRCHIDAMQARVRSVARRHTAARGAVSAGLPVTLDTAAADRRPGSAPTVRAAN
WGLRIAVSTEDLTDPAYHCAVQGHIRKRGVDTCTAADIITDEKRIIVKHLIPO
ALQHTERLKVROVDKWKVTGMGDVCSDFKVPPIHTDGLSTDFWVAVSVSE
DYLAAMATCOVPSDGHPAVGINIPAINIARVYDOLVRYVTHEMAHALGFSDFITD
KMLHNVKTRKRNAPVINSVAARAEQYCDTLEYLEMEDEGCGAGSHIKM
RNAQDELMAPAAAGYSALTMAIFODLGYOAPFKAEVMPGRNAGCAFLSEKME
CGVVPFNGNSCAONAKSMVAOVAEVFVPSDARCTDGFRRKRTETVNTSAGICAM
VLCDTATRTYSVOYGGSGYACTPGILRVELSTVSSAFEBGGYITCPPYVEVCGQNV
1..39
/gene="msp2"
97..1794
/gene="msp2"
/product="glycoprotein 63"

CDS

gene

sig_peptide
mat_peptide
ORIGIN
Query Match
Best Local Similarity 61.2%; Score 29.4; DB 3; Length 1797;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

1 TATGATCAGCTGGTTACCGGCTGTTTACCATGAAATGGACATGC 47
751 TACGACGAGCTGTGACGCTGTCTGACGACGAGATGGCGACGC 797

RESULT 2

LOCUS LIGP63GEN 1860 bp DNA linear INV 16-DEC-1996
DEFINITION L. infantum GP63 gene.
ACCESSION Z83677
VERSION Z83677.1 GI:1743284
KEYWORDS GP63 gene; protease.
SOURCE Leishmania infantum
ORGANISM Leishmania infantum
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1 (bases 1 to 1860)
AUTHORS Ortiz, G., Martinez, P. and Segovia, M.
TITLE Characterization of GP63 of Leishmania infantum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1860)
AUTHORS Ortiz, G.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-1996) Ortiz G., Facultad de Medicina, Universidad
de Murcia, Genetica y Microbiologia, Campus de Espinardo, Murcia,
Murcia, SPAIN, 30100
Location/Qualifiers
1..1860
/organism="Leishmania infantum"

FEATURES

source

CDS

/mol_type="genomic DNA"
/strain="cutaneous"
/db_xref="taxon:5671"
/dev_stage="promastigote"
16..1815
/function="protease"
/codon_start=1
/product="GP63"
/protein_id="CAB06018.1"
/db_xref="GI:1743285"
/translation="MSVSSSTHRRSVARLVRLAAGAIVAAVGTAAAMAHAGV
OHRCHIDAMQARVRSVARRHTAARGAVSAGLPVTLDTAAADRRPGSAPTVRAAN
WGLRIAVSTEDLTDPAYHCAVQGHIRKRGVDTCTAADIITDEKRIIVKHLIPO
ALQHTERLKVROVDKWKVTGMGDVCSDFKVPPIHTDGLSTDFWVAVSVSE
DYLAAMATCOVPSDGHPAVGINIPAINIARVYDOLVRYVTHEMAHALGFSDFITD
KMLHNVKTRKRNAPVINSVAARAEQYCDTLEYLEMEDEGCGAGSHIKM
RNAQDELMAPAAAGYSALTMAIFODLGYOAPFKAEVMPGRNAGCAFLSEKME
CGVVPFNGNSCAONAKSMVAOVAEVFVPSDARCTDGFRRKRTETVNTSAGICAM
VLCDTATRTYSVOYGGSGYACTPGILRVELSTVSSAFEBGGYITCPPYVEVCGQNV
1..39
/gene="msp2"
97..1794
/gene="msp2"
/product="glycoprotein 63"

ORIGIN

Query Match
Best Local Similarity 61.2%; Score 29.4; DB 3; Length 1860;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

1 TATGATCAGCTGGTTACCGGCTGTTTACCATGAAATGGACATGC 47
766 TACGACGAGCTGTGACGCTGTCTGACGACGAGATGGCGACGC 812

RESULT 3

LOCUS LDO495002 2104 bp DNA linear INV 22-JUL-2002
DEFINITION Leishmania donovani mspc gene for GP63, strain
WHOW/KE/0000/Neal-R1.
ACCESSION AU495002
VERSION AU495002.1 GI:21954455
KEYWORDS GP63; major surface protease; mspc gene.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

REFERENCE 1
AUTHORS Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2104)
AUTHORS Mauricio, I.L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM
Location/Qualifiers
1..2104
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="WHOW/KE/0000/Neal-R1"
/db_xref="taxon:5661"
/country="Kenya"
15..1940
/gene="mspc"
15..1940
/gene="mspc"
/note="constitutive major surface protease"
/codon_start=1
/allele="mspcDelta1"
/product="GP63"
/protein_id="CAD42811.1"
/db_xref="GI:21954456"
/db_xref="GOA:O8MNZ5"
/db_xref="SPTRMBL:O8MNZ5"

FEATURES	source
gene	<p>15. .133</p> <p>/gene="mspc"</p> <p>/size=.1823</p> <p>/gene="mspc"</p> <p>/product="GP63"</p>
CDs	<p>15. .133</p> <p>/gene="mspc"</p> <p>/size=.1823</p> <p>/gene="mspc"</p> <p>/product="GP63"</p>
REFERENCE	<p>1 Mauricio, I.L., Stothard, J.R. and Miles, M.A. Genetic diversity in the Leishmania donovani complex Unpublished 2 (bases 1 to 2106)</p> <p>2 Mauricio, I.L. Direct Submission Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM</p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.</p>
Db	<p>765 TACGACGAGCTGTGACGCGTGTGTCGACGACGAGATGGCGACGC 811</p>
Query Match	<p>61.2%; Score 29.4; DB 3; Length 2104;</p> <p>Best Local Similarity 76.6%; Pred. No. 2.8;</p> <p>Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;</p>
Orig	<p>1 TATGATGACGTGTACCGCTGTGTTCATCCATGAAGGACATGC 47</p> <p> </p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.</p>
REFERENCE	<p>1 Mauricio, I.L., Stothard, J.R. and Miles, M.A. Genetic diversity in the Leishmania donovani complex Unpublished 2 (bases 1 to 2106)</p> <p>2 Mauricio, I.L. Direct Submission Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM</p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.</p>
REFERENCE	<p>1 Mauricio, I.L., Stothard, J.R. and Miles, M.A. Genetic diversity in the Leishmania donovani complex Unpublished 2 (bases 1 to 2106)</p> <p>2 Mauricio, I.L. Direct Submission Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM</p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.</p>
REFERENCE	<p>1 Mauricio, I.L., Stothard, J.R. and Miles, M.A. Genetic diversity in the Leishmania donovani complex Unpublished 2 (bases 1 to 2106)</p> <p>2 Mauricio, I.L. Direct Submission Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM</p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.</p>
REFERENCE	<p>1 Mauricio, I.L., Stothard, J.R. and Miles, M.A. Genetic diversity in the Leishmania donovani complex Unpublished 2 (bases 1 to 2106)</p> <p>2 Mauricio, I.L. Direct Submission Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM</p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.</p>
REFERENCE	<p>1 Mauricio, I.L., Stothard, J.R. and Miles, M.A. Genetic diversity in the Leishmania donovani complex Unpublished 2 (bases 1 to 2106)</p> <p>2 Mauricio, I.L. Direct Submission Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM</p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.</p>
REFERENCE	<p>1 Mauricio, I.L., Stothard, J.R. and Miles, M.A. Genetic diversity in the Leishmania donovani complex Unpublished 2 (bases 1 to 2106)</p> <p>2 Mauricio, I.L. Direct Submission Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM</p>
LOCUS	<p>LD0495006 2106 bp DNA linear INV 22-JUL-2002</p> <p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
DEFINITION	<p>Leishmania donovani mspc gene for GP63, strain MCAN/IQ/1977/Bummi3.</p>
ACCESSION	<p>AJ495006</p>
VERSION	<p>AJ495006.1 GI:21954463</p>
KEYWORDS	<p>GP63; major surface protease; mspc gene.</p>
SOURCE	<p>Leishmania donovani</p>
ORGANISM	<p>Leishmania donovani</p> <p>Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomat</p>

Query Match	61.2%	Score 29.4	DB 3	Length 2106
Best Local Similarity	76.6%	Pred. No. 2.8		
Matches	36	Conservative 0	Mismatches 11	Indels 0
Gaps	0			
Db	764	TATGATCAGCTGTTACCCGTTGTTTACCAGTAATGGCAGATGC	47	
LOCUS	LD0495007	LD0495007	2106 bp	DNA linear
DEFINITION	Leishmania donovani mspc gene for GP63, strain MHOM/SU/1984/Marz-Krim.			
ACCESSION	AJ495007			
VERSION	AJ495007.1	GI:21954465		
KEYWORDS	GP63; major surface protease; mspc gene.			
SOURCE	Leishmania donovani			
ORGANISM	Leishmania donovani			
REFERENCE	1	Leishmania		
AUTHORS	Mauricio, I.L., Stothard, J.R. and Miles, M.A.			
TITLE	Genetic diversity in the Leishmania donovani complex			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 2106)			
AUTHORS	Mauricio, I.L.			
TITLE	Direct Submission			
JOURNAL	Submitted (11-JUL-2002)			
FEATURES	Diseases, London School of Hygiene and Trop. Med., Keppel Street, WC1E 7HT, UNITED KINGDOM			
source	Location/Qualifiers			
	1..2106			
	/organism="Leishmania donovani"			
	/mol_type="genomic DNA"			
	/strain="MHOM/SU/1984/Marz-Krim"			
	/db_xref="taxon:5661"			
	14..1939			
	/gene="mspc"			
	14..1939			
	/gene="mspc"			
	/note="constitutive major surface protease"			
	/codon_start=1			
	/allele="mspcldh"			
	/product="GP63"			
	/protein_id="CAD42816.1"			
	/db_xref="GI:21954466"			
	/db_xref="SPTREMBL:O8MM48"			
	/translation="MSVDSSTHRSVSARLVRILAAAGAVTIAVCTAAAHAGAV			
	OHRCTIDAHMPRVQSVARHRTAPGAVSAGLRYTLDTPAADRRPGSAFTVRRAN			
	WGALRIHASTEDLTPAIVHARGVQRNNHAGAIATCTADIDLTDRKDLVLYLIPQ			
	ALDLHERLKRQVODKMKYTDVDEICGFPKPAHITDGLSTDFVMVAAPBEE			
	GVLAATTCQVPSDGHBAVGINI PAANIASRYDQVTRVVTHEMAHAGSGSTFTE			
	ILVVTOMNIRKQDPNVSIVNSSTAVAKAEQYCGGLIETLIEDQGAAGSFKSKE			
	RNKADLMARAAAGYSALTYMILFODLGYQADPSKAEPMGRNAGCAFLSEKME			
	QNTTKPAHFCNVSVDVRCPTRLMLGTCGIGISTPTSPSYQYITNLSLGSYSPL			
	DYCPVLYGDSQNDPASIATGFGAIFNVFSDAACIDGAFRPNKRTAADGYAGIC			
	ANRCDPATATYSVQVCGSMADYVNCPTGRLVELISTVSSAFEEGGYITCPPYVQAN			

sig_peptide
mat_peptide
ORIGIN
VKGADPACDSDSSSSAGDADRAAMORWMDRAGLATAAMVLGMVLVVMVLT
LLITCPMWCCKFGGLPT"
14..130
/gene="mspc"
131..1822
/gene="mspc"
/product="GP63"

Query Match
Best Local Similarity 61.2%; Score 29.4; DB 3; Length 2106;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TATGATGAGCTGGTTACCCGCTGTTTACCCATGAATGCGACATGC 47
764 TACGACCGAGCTGGTACAGCTGTCTGACGACGAGATGGCGCACGC 810

RESULT 6
LDO495005
LOCUS
DEFINITION
LDO495005 2107 bp DNA linear INV 22-JUL-2002
Leishmania donovani mspc gene for GP63, strain
ACCESSION
AJ495005
VERSION
AJ495005.1 GI:21954461
KEYWORDS
GP63; major surface protease; mspc gene.
SOURCE
Leishmania donovani
ORGANISM
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE
AUTHORS
1 Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE
Genetic diversity in the Leishmania donovani complex
JOURNAL
Unpublished
2 (bases 1 to 2107)
REFERENCE
AUTHORS
Mauricio, I.L.
TITLE
Direct Submission
JOURNAL
Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM

FEATURES
source
Location/Qualifiers
1..2107
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MCAN/10/1981/Sukkar2"
/db_xref="taxon:5661"
/country="Iraq"
14..1939
/gene="mspc"
14..1939
/gene="mspc"
/note="constitutive major surface protease"
/codon_start=1
/allele="mspcIdg1"
/product="GP63"
/protein_id="CAD2814.1"
/db_xref="GI:21954462"
/db_xref="GOA:O8WNZ2"
/db_xref="SPTREMBL:O8WNZ2"
/translation="MSVDSSTHRSVAARLVRLAAGAIVAAVGAAMAAGAV
WGLARLAVSTEDITDPAVCAVQVNNHAGALATCAEDILTDERDILVKLLIPQ
ALQHTERLKVROVDKRVNVDVDEICDFVPPQAHITGGSTDTVMVVASPSE
GVLAMATTCQVSDGHPAVGVNIPANIASRYDQLVTRVYTHMAHALGSGFTFE
ILVVTQMNIRKGDENVSVINSTAVAKARQYCGDTLEVTLEDQAGSAGSHIKM
RANKDELMAFAAAGVYSALTMALFODLGFQADPSKAEEMPRNAGCAFLSEKME
QNTTKPMFACNVSDVVRCPTRMLTCCIRGYSSTPSPYQWYFTNISLGGYSPL
DVCPEVIGAGDSNODASLATFGFAGNVNSDARCIDAFPRKRNATADGYAGIC
AVNRCIATRTYSVQVRSMDYNTPELRVLELSTVSSAEFEGYITCPYVEVCOAN
VKGADPAGDSNSSSSAGDADRAAMORWMDRAGLATAAMVLGMVLVVMVLT
LLITCPMWCCKFGGLPT"

gene
CDS

sig_peptide
/gene="mspc"

mat_peptide
ORIGIN
131..1822
/gene="mspc"
/product="GP63"

Query Match
Best Local Similarity 61.2%; Score 29.4; DB 3; Length 2107;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TATGATGAGCTGGTTACCCGCTGTTTACCCATGAATGCGACATGC 47
764 TACGACCGAGCTGGTACAGCTGTCTGACGACGAGATGGCGCACGC 810

RESULT 7
LDO495003
LOCUS
DEFINITION
LDO495003 2109 bp DNA linear INV 22-JUL-2002
Leishmania donovani mspc gene for GP63, strain
ACCESSION
AJ495003
VERSION
AJ495003.1 GI:21954457
KEYWORDS
GP63; major surface protease; mspc gene.
SOURCE
Leishmania donovani
ORGANISM
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE
AUTHORS
1 Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE
Genetic diversity in the Leishmania donovani complex
JOURNAL
Unpublished
2 (bases 1 to 2109)
REFERENCE
AUTHORS
Mauricio, I.L.
TITLE
Direct Submission
JOURNAL
Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM

FEATURES
source
Location/Qualifiers
1..2109
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MHOM/KE/1954/LRC-L53"
/db_xref="taxon:5661"
/country="Kenya"
15..1940
/gene="mspc"
15..1940
/gene="mspc"
/note="constitutive major surface protease"
/codon_start=1
/allele="mspcIda2"
/product="GP63"
/protein_id="CAD2812.1"
/db_xref="GI:21954458"
/db_xref="GOA:O8WNZ4"
/db_xref="SPTREMBL:O8WNZ4"
/translation="MSVDSSTHRSVAARLVRLAAGAIVAAVGAAMAAGAV
WGLARLAVSTEDITDPAVCAVQVNNHAGALATCAEDILTDERDILVKLLIPQ
ALQHTERLKVROVDKRVNVDVDEICDFVPPQAHITGGSTDTVMVVASPSE
GVLAMATTCQVSDGHPAVGVNIPANIASRYDQLVTRVYTHMAHALGSGFTFE
ILVVTQMNIRKGDENVSVINSTAVAKARQYCGDTLEVTLEDQAGSAGSHIKM
RANKDELMAFAAAGVYSALTMALFODLGFQADPSKAEEMPRNAGCAFLSEKME
QNTTKPMFACNVSDVVRCPTRMLTCCIRGYSSTPSPYQWYFTNISLGGYSPL
DVCPEVIGAGDSNODASLATFGFAGNVNSDARCIDAFPRKRNATADGYAGIC
AVNRCIATRTYSVQVRSMDYNTPELRVLELSTVSSAEFEGYITCPYVEVCOAN
VKGADPAGDSNSSSSAGDADRAAMORWMDRAGLATAAMVLGMVLVVMVLT
LLITCPMWCCKFGGLPT"

gene
CDS

sig_peptide
mat_peptide
ORIGIN
/gene="mspc"
/product="GP63"

Query Match 61.2%; Score 29.4; DB 3; Length 2109;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGGTTACCGGTGTGTTTACCCATGAATGGACATGC 47
765 TACGACGAGCTGTGACGCGTGTCTCAGCAGCAGATGGCGCAGC 811

RESULT 8
LDO495004 2109 bp DNA linear INV 22-JUL-2002
LOCUS Leishmania donovani mspc gene for GP63, strain
DEFINITION MHOM/CN/0000/Mang1e1.
ACCESSION AJ2495004.1 GI:21954459
VERSION GP63; major surface protease; mspc gene.
KEYWORDS Leishmania donovani
SOURCE Leishmania donovani
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1
AUTHORS Mauricio, I.L., Storchard, J.R. and Miles, M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2109)
AUTHORS Mauricio, I.L.
TITLE Direct Submission
SUBMITTED (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM

FEATURES
source location/Qualifiers
1..2109
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MHOM/CN/0000/Mang1e1"
/db_xref="taxon:5661"
/country="China"
15..1940
/gene="mspc"
15..1940
/note="constitutive major surface protease"
/codon_start=1
/allele="mspcLdF"
/product="GP63"
/protein_id="CAD2813.1"
/db_xref="GI:21954460"
/db_xref="GOA:O8MN23"
/db_xref="SPTR:EMBL.O8MN23"
/translation="MSVDSSTTRHRSVARRLRLAAGAATVIAVGTAAAMAHAGAV
OHRCTIDAMQARVQSVARRHTAPGAVSAGLPTVLTDTAAADRSGAPVVRAN
WGLRLAVSTEDLTDPAYHCAVGOHVNNAAGATCTABDILTDEKRDILVGLILPQ
ALQHTERLKVQVQDKMKTVMDEICDPKVPQAHITGFSNTDFVMVVASVSE
GVLAAMTTCCVPSDGHAPVAVINI PAANIASRYDQALVTRVTHMAHALGFSVGFPE
ILVTCOMNIRGDFDPVPTVNSGRAVAKARBOVCDLTLEIETEDOGASASGSHITM
RNAQDELMAAPAAAGYSALTMAIFDQLGTYQADFSAEVMPGRNAGAFLEKCKE
RNTTKMPAMFCNESEVYTRMCPSTRLSLIGKGVYRHPDLPTWQYFTDPSLAGISAFMD
DYCEPVIYSGDSQNDASLAAGFFSAFNVSFSDARCIDGAFRKNRTAANGYAGLC
ANVCDDATRTYSQVHGSGMYNCTGLRVELSTVSFAFEEGGYITCPPYVEVCOGN
VOAKDPAGDSSSSAGDAADRAAQRMDRMAGLATAAMVLLGWLISLMLVYVWL
LLLTCPWVCKCFGLPFP"

ORIGIN
819_peptide 15..131
/gene="mspc"
mat_peptide 132..1823
/gene="mspc"
/product="GP63"

Query Match 61.2%; Score 29.4; DB 3; Length 2109;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Query Match 61.2%; Score 29.4; DB 3; Length 2892;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGGTTACCGGTGTGTTTACCCATGAATGGACATGC 47
765 TACGACGAGCTGTGACGCGTGTCTCAGCAGCAGATGGCGCAGC 811

RESULT 9
LEIGP63D 2892 bp mRNA linear INV 26-APR-1993
LOCUS Leishmania donovani chagasi surface protease (log glycoprotein 63,
DEFINITION GP63) mRNA, complete cds.
ACCESSION M80672.1 GI:159328
VERSION M80672.1 GI:159328
KEYWORDS log glycoprotein 63; surface protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1 (bases 1 to 2892)
AUTHORS Ramamoorthy, R., Donelson, J.E., Paetz, K.E., Maybodi, M., Roberts, S.C.
and Wilson, M.E.
TITLE Three distinct RNAs for the surface protease gp63 are
JOURNAL differentially expressed during development of Leishmania donovani
MEDLINE J. Biol. Chem. 267 (3), 1888-1895 (1992)
PUBMED 92112918
1370484
COMMENT Original source text: Leishmania donovani (sub_species chagasi)
cDNA to mRNA.

FEATURES
source location/Qualifiers
1..2892
/organism="Leishmania donovani"
/mol_type="mRNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
1..2892
/gene="GP63"
114..1913
/note="log glycoprotein 63"
/codon_start=1
/product="surface protease"
/protein_id="AAA29238.1"
/db_xref="GI:159328"
/translation="MSVDSSTTRHRSVARRLRLAAGAATVIAVGTAAAMAHAGAV
OHRCTIDAMQARVQSVARRHTAPGAVSAGLPTVLTDTAAADRSGAPVVRAN
WGLRLAVSTEDLTDPAYHCAVGOHVNNAAGATCTABDILTDEKRDILVGLILPQ
ALQHTERLKVQVQDKMKTVMDEICDPKVPQAHITGFSNTDFVMVVASVSE
GVLAAMTTCCVPSDGHAPVAVINI PAANIASRYDQALVTRVTHMAHALGFSVGFPE
ARLIESISNVRHDDPVPVPTVNSGRAVAKARBOVCDLTLEIETEDOGASASGSHITM
RNAQDELMAAPAAAGYSALTMAIFDQLGTYQADFSAEVMPGRNAGAFLEKCKE
RNTTKMPAMFCNESEVYTRMCPSTRLSLIGKGVYRHPDLPTWQYFTDPSLAGISAFMD
DYCEPVIYSGDSQNDASLAAGFFSAFNVSFSDARCIDGAFRKNRTAANGYAGLC
ANVCDDATRTYSQVHGSGMYNCTGLRVELSTVSFAFEEGGYITCPPYVEVCOGN
VOAKDPAGDSSSSAGDAADRAAQRMDRMAGLATAAMVLLGWLISLMLVYVWL
LLLTCPWVCKCFGLPFP"

ORIGIN
819_peptide 114..230
/gene="GP63"
mat_peptide 405..1910
/gene="GP63"
polya_site 2892
/gene="GP63"

Query Match 61.2%; Score 29.4; DB 3; Length 2892;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGGTTACCGGTGTGTTTACCCATGAATGGACATGC 47
864 TACGACGAGCTGTGACGCGTGTCTCAGCAGCAGATGGCGCAGC 910

ACCESSION	M28527
VERSION	M28527.1 GI:159322
KEYWORDS	glycoprotein; protease; surface antigen.
SOURCE	Leishmania donovani
ORGANISM	Leishmania donovani Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE	1 (bases 1 to 3047)
AUTHORS	Miller, R.A., Reed, S.G. and Parsons, M.
TITLE	Leishmania gp63 molecule implicated in cellular adhesion lacks an Arg-Gly-Asp sequence
JOURNAL	Mol. Biochem. Parasitol. 39 (2), 267-274 (1990)
MEDLINE	90205976
PUBMED	2320059
COMMENT	Original source text: L.chagasi (isolate MHOM/BR/82/BA-2Cia) DNA, clones pLc63-1 and 2).
FEATURES	Defect entry and computer-readable sequence for [1] kindly provided by M.Parsons, 03-OCT-1989.
SOURCE	Location/Qualifiers .3047 /organism="Leishmania donovani" /mol_type="genomic DNA" /sub_species="chagasi" /db_xref="taxon:5661" 496..2295 /note="gp63 protein" /codon_start=1 /protein_id="AA92935.1" /db_xref="GI:159323" /translation="MSVDSSTHRRSVARLVRLAAGAAYIAVAIVRAAHAGAT OHCICIDAMQARVROGVARHHTAPGAVSAGEPYUTLTPTAADRPSPALFVVAAL WGLRIASTEDTDPAYHCARGCHIRKELGCVITCTREDILTDEKDILVKHLIP ALOTHEERLKRVONODKKVTMGSDPFCVPAPHITDGLSNTDFVMYVASVSEI GYLAATTCGVPSDSHPAVGVINSTITPAANIASRYDOLVTVVTHEMAHALGSVFPEE AAIHLSISVRHKDDPVDVINSTIAVAAREYGCDLEYLEIEDOGAGSGASHIKR RNAHDPMAFNPAAGAYSAITMAIFQDIGFYQADSKAENVMPGNAGCAFLSECKM KNITTPMAPFNENEVTRCPSTRSLSKCGVTRHPDLPIYMOWFTDTSLAGISFPM CCPVVEPYDGSCAQRSBESAGAPFKGFNVFSPARCTDGAERPRTSHIITSYGLCLC NCDDPATRYTYSVOVGHSYGVAICTPGIRVELSYTSSAFEEGGYTTCPPYEVCQNAV QAARKGNAAGARRGRPRAAATPALVALAIALAL"
ORIGIN	
Query Match	61.2%; Score 29.4; DB 3; Length 3047;
Best Local Similarity	76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;	
Dn	1246 TACGACCAGCTGTGTGACGCCGTGTGTGCACGACGATGGCGCACGC 1292
Cy	1 TATGATCAGCTGTGTTACCCGCTGTTGTTCCCATGAATAAGCACATGC 47
RESULT 12	
LOCUS	L1U48798 3047 bp DNA linear INV 15-JAN-1998
DEFINITION	Leishmania infantum surface glycoprotein gp63 (gp63) gene, complete cds.
ACCESSION	U48798
VERSION	U48798.1 GI:1213329
KEYWORDS	
SOURCE	. Leishmania infantum Leishmania infantum Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
ORGANISM	1 (bases 1 to 3047) Gonzalez-Aleguino-laza, G., Almazan, F., Rodriguez, J.F., Marquet, A. and Larrea, V.
REFERENCE	Cloning of the gp63 surface protease of Leishmania infantum. different post-translational modifications correlated with different infective forms
AUTHORS	Biochim. Biophys. Acta 1361 (1), 92-102 (1997)
TITLE	
JOURNAL	
MEDLINE	97388259
PUBMED	9247093
REFERENCE	2 (bases 1 to 3047)

AUTHORS Gonzalez-Abeguinolaza, G., Almazan, F., Rodriguez-Aguirre, J.F., Marquet, A. and Larraga, V.

TITLE Direct Submission

JOURNAL Submitted (08-FEB-1996) Vicente Larraga, Proteins Structure, Center of Biological Research, Spanish Research Council, Velazquez 144, Madrid 28006, Spain

FEATURES

SOURCE

1. 3047
/organism="Leishmania infantum"
/mol_type="genomic DNA"
/db_xref="taxon:5671"
496.2295
/gene="gp63"
496.2295
/gene="gp63"
/note="protease"
/codon_start=1
/product="surface glycoprotein gp63"
/protein_id="AAB96339.1"
/db_xref="GI:1213330"

gene

CDs

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 3047;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

1 TATGATCAGCTGTTACCGGTGTTTACCCATGAATGGACATGC 47
1246 TACGACGCTGTTACCGGTGTTTACCCATGAATGGACATGC 1292

RESULT 13

LEIGPAA 3105 bp DNA linear INV 26-APR-1993

LOCUS L.donovani.

ACCESSION M60048

VERSION M60048.1 GI:159334

KEYWORDS glycoprotein 63.

SOURCE Leishmania donovani

ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

REFERENCE 1 (bases 1 to 3105)
AUTHORS Webb, J.R., Button, L.L. and McMaster, W.R.
TITLE Heterogeneity of the genes encoding the major surface glycoprotein of Leishmania donovani

JOURNAL Mol. Biochem. Parasitol. 48 (2), 173-184 (1991)

MEDLINE 92107220

PUBMED 1762629

COMMENT Original

FEATURES source text: L.donovani DNA.
Location/Qualifiers
1. 3105
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/db_xref="taxon:5661"
101.1873
/gene="gp63"
101.1873
/gene="gp63"
/codon_start=1
/product="glycoprotein 63"
/protein_id="AAA29244.1"

/db_xref="GI:159335"
/translation="MSVDSSTHRRHSVAARLVRLAAGAATVTAAGTAAMAAGAV
OHRCHIDAMQARVRSVARRHTAPGASAVGLPYTLDTPAAADRPGSAFTVRAAN
EDLDTSAVHCAVQRISTRDGRPAITCAEDILTDKRDILVKLIPQALQHTERLK
VRQVQDMKATGKMGNEICGHPKVPPIHTGLSTNTPVMVAASVSSGDVLAAMATTCQ
VFSQHPAVAVINI PAANIASRYDQVTRVTHMAALGFSVGFPEG
RHKDFVINSSTPAVAKARQYCGTLETEMDOGASGSHIKRNADODLISNV
ASADGVYSAITMAI FODLGFYQADPSKAEPMPCRNACGAFISKCMEDGTTKPMF
CNEBETVTRCHTGRSLGVCGLSSDIPLPYMOYFPDPLLAGISAFMDYCPVVPFG
DGSCAORASEAGAPFKGNVSDARCTIDAFRKTETVNSYAGICANVCDTAR
TVSVQHGSGSYANCTPGLRVELSTVSSAFEBEGYITCPPYVEVCOGNVQAAKDGNA
AAGRGRPAATALLVALAVAL"

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 3105;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

1 TATGATCAGCTGTTACCGGTGTTTACCCATGAATGGACATGC 47
821 TACGACGCTGTTACCGGTGTTTACCCATGAATGGACATGC 867

RESULT 14

LEIGP63C 2621 bp mRNA linear INV 26-APR-1993

LOCUS Leishmania donovani chagasi surface protease (constitutive glycoprotein 63, GP63) mRNA, complete cds.

ACCESSION M80671

VERSION M80671.1 GI:159326

KEYWORDS constitutive glycoprotein 63; surface protease.

SOURCE Leishmania donovani

ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

REFERENCE 1 (bases 1 to 2621)
AUTHORS Ramamoorthy, R., Donelson, J.E., Paetz, K.E., Maybodi, M., Roberts, S.C. and Wilson, M.E.
TITLE Three distinct RNAs for the surface protease gp63 are differentially expressed during development of Leishmania donovani chagasi promastigotes to an infectious form

JOURNAL J. Biol. Chem. 267 (3), 1888-1895 (1992)

MEDLINE 92112918

PUBMED 1370484

COMMENT Original source text: Leishmania donovani (sub-species chagasi) CDNA to mRNA.
Location/Qualifiers
1. 2621
/organism="Leishmania donovani"
/mol_type="mRNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
1. 2621
/gene="GP63"
132.2051
/gene="GP63"
/note="constitutive glycoprotein 63"
/codon_start=1
/product="surface protease"
/protein_id="AAA29237.1"
/db_xref="GI:159327"

FEATURES

SOURCE source text: "Leishmania donovani"
Location/Qualifiers
1. 2621
/organism="Leishmania donovani"
/mol_type="mRNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
1. 2621
/gene="GP63"
132.2051
/gene="GP63"
/note="constitutive glycoprotein 63"
/codon_start=1
/product="surface protease"
/protein_id="AAA29237.1"
/db_xref="GI:159327"

/translation="MSVDSSTHRRHSVAARLVRLAAGAATVTAAGTAAMAAGAV
OHRCHIDAMQARVRSVARRHTAPGASAVGLPYTLDTPAAADRPGSAFTVRAAN
EDLDTSAVHCAVQRISTRDGRPAITCAEDILTDKRDILVKLIPQALQHTERLK
VRQVQDMKATGKMGNEICGHPKVPPIHTGLSTNTPVMVAASVSSGDVLAAMATTCQ
VFSQHPAVAVINI PAANIASRYDQVTRVTHMAALGFSVGFPEG
RHKDFVINSSTPAVAKARQYCGTLETEMDOGASGSHIKRNADODLISNV
ASADGVYSAITMAI FODLGFYQADPSKAEPMPCRNACGAFISKCMEDGTTKPMF
CNEBETVTRCHTGRSLGVCGLSSDIPLPYMOYFPDPLLAGISAFMDYCPVVPFG
DGSCAORASEAGAPFKGNVSDARCTIDAFRKTETVNSYAGICANVCDTAR
TVSVQHGSGSYANCTPGLRVELSTVSSAFEBEGYITCPPYVEVCOGNVQAAKDGNA
AAGRGRPAATALLVALAVAL"

sig_peptide 132..248
/gene="GP63"
mat_peptide 423..2048
/gene="GP63"
/product="surface protease"
/note="constitutive glycoprotein 63"
2621
/gene="GP63"
ORIGIN
polyA_site
Query Match 59.2%; Score 28.4; DB 3; Length 2621;
Best Local Similarity 76.1%; Pred. No. 6.8;
Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Oy 1 TATGATCAGCTGGTTACCGCTGTTTACCCATGAATGGCACANG 46
Db 882 TAGACACAGCTGTGACGCGTGTCTGACGACGAGATGGCCACAG 927
RESULT 15
LOCUS LTU495010 2050 bp DNA linear INV 22-JUL-2002
DEFINITION Leishmania turanica mspc gene for GP63, strain
MRHO/SU/1983/MARZ-051.
ACCESSION AJ495010
VERSION AJ495010.1 GI:21954471
KEYWORDS GP63; major surface protease; mspc gene.
SOURCE Leishmania turanica
ORGANISM Leishmania turanica
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1
AUTHORS Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2050)
AUTHORS Mauricio, I.L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM
FEATURES
source Location/Qualifiers
1. 2050
/organism="Leishmania turanica"
/mol_type="genomic DNA"
/strain="MRHO/SU/1983/MARZ-051"
/db_xref="taxon:62297"
15..1949
/gene="mspc"
15..1949
/gene="mspc"
/note="constitutive major surface protease"
/codon_start=1
/allele="mspc.tu1"
/product="GP63"
/protein_id="CAD42819.1"
/db_xref="GI:21954472"
/db_xref="GOA:Q8MNY9"
/db_xref="SPTREMBL:Q8MNY9"
/translation="MSVDSSTHRRRCVAAIRLVRLAAGVAVTVAGTAAAHAGAL
OHRCHDAMQARVROSVDHHTAPGAVSAVG.PYVTLDAHTAAADPRGARGVER
DVWNGALRTAIEEDLTDPYRHCVRGQVNNHAGAVITCTABDILTDEKRDILVHL
IPRALKHTERLKVQVQGTWKTGVNDEICDFFKPOAHITKFSNTDFWYVASVP
SERGVLAWATTCOVFSBGRPAVGVINIPANIVSRYDQLVTRVTHEMAHALGFSGP
FEDARIVANISVGRNRPVPIVNSSTAVAKAREYGCDTLEYLEVEDQAGSAGSH
IKRNARODELMAPAAAGYTLALMALFODLGFYQADPSKAEVLMGRNAGCALPTDK
CMEONTOWPAMCNESEDAIRCPSTRILIGCGIREVETPIPIYQVPTNRTASLGYS
PFLDYCPSVIGVAHGKCNODASAAELIAAFVPSPAKCIDGATTPKRTANSGYS
GLCANVQCDTARTTYSVQVHGSMDYANCTPGRLVELSTVSNTEEGYITCEPYEVC
QGNVKGKDFDGSDDSSSAGDDADKAAMERNRMDWAGLATAVTVLIGIVLSIMALV
VWLLIVSCPRWCCKVGLPT"
15..131
/gene="mspc"
sig_peptide

mat_peptide 132..1832
/gene="mspc"
/product="GP63"
ORIGIN
Query Match 57.9%; Score 27.8; DB 3; Length 2050;
Best Local Similarity 74.5%; Pred. No. 11;
Matches 35; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Oy 1 TATGATCAGCTGGTTACCGCTGTTTACCCATGAATGGCACATGC 47
Db 774 TAGACACAGCTGTGACGCGTGTCTGACGACGAGATGGCCACGC 820
Search completed: March 15, 2004, 22:49:58
Job time : 210.439 secs

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fibrillar presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fibrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fibrillar subunit proteins are usually strong
CC against the inserted epitope, and hybrid fibrillae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 48 BP; 12 A; 11 C; 11 G; 14 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.4e-10;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGACATGCA 48
1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGACATGCA 48

RESULT 2

AAC64609/c
ID AAC64609 standard; DNA; 78 BP.

AC AAC64609;

DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)

DE Agfa (SEF17) recombinant agfa::PT3a generating PCR primer 17-B.
KW Salmonella; agfa; chromosomal gene replacement; fimbtrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.

OS Salmonella enteritidis.
OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Example 2; Page 61; 139p; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Agfa, CsgA and Agfa-homologue fimbtrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fibrillar presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fibrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fibrillar subunit proteins are usually strong
CC against the inserted epitope, and hybrid fibrillae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)

Sequence 78 BP; 17 A; 17 C; 21 G; 23 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 9.5e-10;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGACATGCA 48
48 TATGATCAGCTGTTACCCGTTGTTACCATGAATGACATGCA 1

RESULT 3

AAC64610
ID AAC64610 standard; DNA; 78 BP.

AC AAC64610;

DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)

DE Agfa (SEF17) recombinant agfa::PT3a generating PCR primer 17-C.
KW Salmonella; agfa; chromosomal gene replacement; fimbtrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.

OS Salmonella enteritidis.
OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Example 2; Page 61; 139p; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct US field.)

CC Sequence 78 BP; 21 A; 19 C; 17 G; 21 T; 0 U; 0 Other;

CC Query Match 100.0%; Score 48; DB 3; Length 78;

CC Best Local Similarity 100.0%; Pred. No. 9.5e-10; Mismatches 0; Indels 0; Gaps 0;

CC Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC 1 TATGATCAGCTGTTACCCGTTGTTTACCCATGAATGCAATGCA 48
CC 1 TATGATCAGCTGTTTACCCGTTGTTTACCCATGAATGCAATGCA 48

CC RESULT 4
CC AAC64628
CC ID AAC64628 standard; DNA; 456 BP.

CC AC AAC64628;

CC DT 26-FEB-2001 (first entry)

CC DE Agfa::PT3#7 DNA sequence SEQ ID NO:23.

CC KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

CC KM vaccine; immune response; immunogen; ds.

CC OS *Salmonella enteritidis*.

CC OS *Escherichia coli*.

CC OS Synthetic.

CC PN WO200060102-A2.

CC PD 12-OCT-2000.

CC PF 05-APR-2000; 2000WO-CA000356.

CC PR 05-APR-1999; 99US-0127888P.

CC PA (UTVI-) UNIV VICTORIA.

CC PI White AP, Doran JL, Collison SK, Kay WW;

CC DR WPI; 2000-672631/65.

CC DR P-PSDB; AAB36352.

CC PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
CC PT which encodes foreign epitope or antigen, expresses recombinant Agfa
CC PT protein useful for eliciting immune response in animal.

CC PS Disclosure; Page 137; 139pp; English.

CC XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEFI/TF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

CC Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

CC Query Match 100.0%; Score 48; DB 3; Length 456;

CC Best Local Similarity 100.0%; Pred. No. 1.5e-09; Mismatches 0; Indels 0; Gaps 0;

CC Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC 1 TATGATCAGCTGTTACCCGTTGTTTACCCATGAATGCAATGCA 48
CC 331 TATGATCAGCTGTTACCCGTTGTTTACCCATGAATGCAATGCA 378

CC RESULT 5
CC AAC64622
CC ID AAC64622 standard; DNA; 456 BP.

CC AC AAC64622;

CC DT 26-FEB-2001 (first entry)

CC DE Agfa::PT3#1 DNA sequence SEQ ID NO:11.

CC KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;

CC KM vaccine; immune response; immunogen; ds.

CC OS *Salmonella enteritidis*.

CC OS *Escherichia coli*.

CC OS Synthetic.

CC PN WO200060102-A2.

CC PD 12-OCT-2000.

CC PF 05-APR-2000; 2000WO-CA000356.

CC PR 05-APR-1999; 99US-0127888P.

CC PA (UTVI-) UNIV VICTORIA.

CC PI White AP, Doran JL, Collison SK, Kay WW;

CC DR WPI; 2000-672631/65.

CC DR P-PSDB; AAB36346.

CC PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
CC PT which encodes foreign epitope or antigen, expresses recombinant Agfa
CC PT protein useful for eliciting immune response in animal.

DR P-PSDB; AAB36354.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TATGATCAGCTGTTACCGTGTGTTTACCCATGAATGCGACATGCA 48
Db 241 TATGATCAGCTGTTACCGTGTGTTTACCCATGAATGCGACATGCA 288
XX
RESULT 8
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 DNA sequence SEQ ID NO:25.
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; immunogen; ds.
XX
XX *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.
XX
PN MO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UTVI-) UNIV VICTORIA.
XX

PI White AP, Doran JL, Collison SK, Kay MW;
XX
DR WPI; 2000-672631/65.
XX
DR P-PSDB; AAB36353.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TATGATCAGCTGTTACCGTGTGTTTACCCATGAATGCGACATGCA 48
Db 172 TATGATCAGCTGTTACCGTGTGTTTACCCATGAATGCGACATGCA 219
XX
RESULT 9
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#4 DNA sequence SEQ ID NO:17.
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; immunogen; ds.
XX
XX *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.
XX
PN MO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX

XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WM,
 PI MPI, 2000-672631/65.
 XX P-PSDB; AAB36349.
 DR
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS
 XX Disclosure; Page 136; 139pp; English.
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombination of the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;
 Query Match 100.0%; Score 48; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGTACCCGTTGTTTACCATGAATGACATGCA 48
 DB 127 TATGATCAGCTGTACCCGTTGTTTACCATGAATGACATGCA 174
 RESULT 10
 AAC64631
 ID AAC64631 standard; DNA; 456 BP.
 AC AAC64631;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#10 DNA sequence SEQ ID NO:29.
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN W0200060102-A2.
 XX 12-OCT-2000.
 PD
 XX

PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WM,
 PI MPI, 2000-672631/65.
 XX P-PSDB; AAB36355.
 DR
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS
 XX Disclosure; Page 138; 139pp; English.
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombination of the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;
 Query Match 100.0%; Score 48; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGTACCCGTTGTTTACCATGAATGACATGCA 48
 DB 307 TATGATCAGCTGTACCCGTTGTTTACCATGAATGACATGCA 354
 RESULT 11
 AAC64623
 ID AAC64623 standard; DNA; 456 BP.
 AC AAC64623;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#2 DNA sequence SEQ ID NO:13.
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN W0200060102-A2.
 XX

PD		12-OCT-2000.
XX	PF	05-APR-2000; 2000MO-CA000356.
XX	PR	05-APR-1999; 99US-0127688P.
XX	PA	(UUVI-) UNIV VICTORIA.
XX	P1	White AP, Doran JL, Collison SK, Kay WW;
DR	WPI; 2000-672631/65.	
DR	P-PsDB; AAB36347.	
FT		Recombinant agfa gene having a segment replaced by a foreign DNA sequence PT which encodes foreign epitope or antigen, expresses recombinant Agfa XX protein useful for eliciting immune response in animal. XX PS Disclosure; Page 136; 139pp; English.
CC		The present invention describes a recombinant agfa gene (1) where a CC segment of the gene has been replaced by a segment of a foreign DNA CC sequence which encodes a foreign epitope or antigen. Also described are: CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended CC assembly system of strains of Salmonella, Escherichia coli and CC Enterobacteriaceae for the production of fimbriae comprising recombinant CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) CC directing recombination of a recombinant gene into the chromosome of the CC homologous species; (3) directing recombination of a recombinant gene CC back into the chromosome of the homologous species, replacing the native CC copy of that gene; and (4) eliciting an immune response in an animal, CC comprising separating an amino acid polymer comprising a recombinant Agfa CC protein containing a replacement segment or segments of foreign amino CC acid sequence or sequences grown on a Salmonella, E. coli or CC Enterobacteriaceae host cell, from the host cell and introducing the CC polymer into the animal in conjunction with a carrier or diluent. (1) is CC useful for the expression of recombinant Agfa protein which is useful for CC eliciting an immune response in an animal. In a fimbrial presentation CC system the heterologous antigens are presented in high numbers (up to CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the CC immunogenicity and adhesion properties relevant for an efficient live CC vaccine, the carrier fimbrial subunit proteins are usually strong CC immunogens, which may be important for directing an immune response CC against the inserted epitope, and hybrid fimbriae are easy and CC inexpensive to purify in large amount. The present sequence is given in CC the exemplification of the present invention XX SO Sequence 456 BP, 118 A, 109 C, 121 G, 108 T, 0 U, 0 Other;
OY		Query Match 100.0%; Score 48; DB 3; Length 456; Best Local Similarity 100.0%; Pred. No. 1.5e+09; Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0
Dn		1 TATGATCAGCGTGTACCCGGTTGTTCACCATGAATGCGACATGCA 48 352 TATGATCAGCTGTGTTACCCTGGTGTTCACCATGAATGCGACATGCA 399
RESULT 12		
ID	AAC64627	standard; DNA; 456 BP.
XX	AAC64627;	
DT	26-FEB-2001	(first entry)
DE	Agfa::PT#6	DNA sequence SEQ ID NO:21.
KV	Salmonella; agfa;	chromosomal gene replacement; fimbrian; epitope;
OS	vaccine; immune response; immunogen; ds.	
OS	Salmonella enteritidis.	
OS	Escherichia coli.	

OS	Synthetic.
XX	
PX	WO200060102-A2.
XX	
PD	12-OCT-2000.
XX	
PF	05-APR-2000; 2000WO-CAN00356.
XX	
PR	05-APR-1999; 99US-0127888P.
XX	
PA	(UYVI-) UNIV VICTORIA.
PI	White AP, Doran JL, Collison SK, Kay W;
XX	
DR	WPI; 2000-672631/65.
DR	P-PsDB; AAB36351.
XX	
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.
PS	Disclosure; Page 137; 139pp; English.
CC	The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are:
CC	(1) use of thin aggregative fimbriae (SBE17/TAF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant AgfA, CsgA and AgfA-homologue fimbrin subunits, respectively, (2)
CC	directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal,
CC	comprising separating an amino acid polymer comprising a recombinant AgfA protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention
XX	
SO	Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;
Query Match	100.0%; Score 48; DB 3; Length 456;
Best Local Similarity	100.0%; Pred. NO. 1.5E+09;
Matches 48; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	1 TATGATCAGTGTACCGTGTTGGTAACCATGAATAATGCACATGCA 48 Db 262 TATGATCAGTGTGTTACCCTGTTGTTCACCATGAATTGCCATGCA 309
RESULT 13	
AAC64624	ID AAC64624 standard; DNA; 456 BP.
XX	
AC	AAC64624;
XX	
DT	26-FEB-2001 (first entry)
DE	Agfa::PT#3 DNA sequence SEQ ID NO:15.
XX	
XX	Salmonella; agfa; chromosomal gene replacement; fimbrian; epitope; vaccine; immune response; immunogen; ds. KW

XX OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 XX Synthetic.
 XX MO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UUYI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay MW;
 DR WPI; 2000-672631/65.
 XX P-PSDB; AAB36348.
 XX
 PS Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;
 Query Match 100.0%; Score 48; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGCTTACCGTGTGTACCATGAATGCGACATGCA 48
 DB 67 TATGATCAGCTGCTTACCGTGTGTACCATGAATGCGACATGCA 114
 RESULT 14
 AAC64616 standard; DNA; 48 BP.
 XX
 AC AAC64616;
 XX
 DT 06-AUG-2003 (revised)
 DT 26-FEB-2001 (first entry)
 XX

DE S. enteritidis recombinant agfa gene sequence.
 XX
 KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS Synthetic.
 XX MO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UUYI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay MW;
 DR WPI; 2000-672631/65.
 XX
 PS Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Fig 1c; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention. (updated on 06-AUG-2003 to
 CC correct OS field.)
 XX
 SQ Sequence 48 BP; 14 A; 10 C; 10 G; 14 T; 0 U; 0 Other;
 Query Match 90.0%; Score 43.2; DB 3; Length 48;
 Best Local Similarity 93.8%; Pred. No. 7.6e-08;
 Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGCTTACCGTGTGTACCATGAATGCGACATGCA 48
 DB 1 TATGATCAGCTGCTTACCGTGTGTACCATGAATGCGACATGCA 48
 RESULT 15
 AAC64606 standard; DNA; 78 BP.
 XX
 AC AAC64606;
 XX

DT 06-AUG-2003 (revised)
 DT 26-FEB-2001 (first entry)
 DE Sefa (SEF14) recombinant sefa::PT3 generating PCR primer 14-C.
 XX
 XX Salmomella: agfa; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen; PCR primer; 88.
 XX
 OS Salmomella enteritidis.
 OS Synthetic.
 XX
 XX WO20060102-A2.
 XX
 XX 12-OCT-2000.
 PD
 XX 05-APR-2000; 2000WO-CA000356.
 PP
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay W;
 P1
 XX WPI: 2000-672631/65.
 DR
 XX
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS
 XX
 XX Example 2; Page 61; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of Salmomella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmomella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
 CC correct OS field.)
 CC
 XX

Sequence 78 BP; 16 A; 16 C; 21 G; 25 T; 0 U; 0 Other;

Query March 84.6%; Score 40.6; DB 3; Length 78;
 Best local similarity 91.5%; Pred. No. 1e-06;
 Matches 43; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TATGATCAGCTGTTACCGTGTGTACCATGAATGACATGC 47
 |||||
 DB 1 TATGATCAGCTGTTACCGTGTGTACCATGAATGACATGC 47

Search completed: March 15, 2004, 17:51:54
 Job time : 29.8457 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 15, 2004, 16:56:17 ; Search time 5.01734 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48

Sequence: 1 tatgatcagctggtaccg.....cccatgaatgacacatgca 48

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/2/ina/PCUTS_COMB.seq:*

6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	24.6	51.3	3636	US-08-961-527-78	Sequence 78, Appl
C 2	22.4	46.7	1014	US-09-325-932A-15	Sequence 15, Appl
C 3	22	45.8	5406	US-08-961-527-166	Sequence 166, App
C 4	21.8	45.4	7754	US-09-634-238-3	Sequence 3, Appli
C 5	21.4	44.6	648	US-09-252-991A-11586	Sequence 11586, A
C 6	21.4	44.6	963	US-09-252-991A-11645	Sequence 11645, A
C 7	21.4	44.6	987	US-09-252-991A-11794	Sequence 11794, A
C 8	21.4	44.6	1983	US-09-453-702B-36	Sequence 36, Appl
C 9	21.4	44.6	2166	US-08-317-401E-3	Sequence 3, Appli
C 10	21.4	44.6	2877	US-08-317-401E-1	Sequence 1, Appli
C 11	21.4	44.6	9541	US-08-961-527-132	Sequence 132, App
C 12	21	43.8	2975	US-09-543-681A-3167	Sequence 3167, Ap
C 13	21	43.8	6587	US-08-879-337-10	Sequence 10, Appl
C 14	21	43.8	100848	US-08-879-337-3	Sequence 3, Appli
C 15	21	43.8	100848	US-09-596-002-39	Sequence 39, Appl
C 16	20.8	43.3	705	US-09-543-681A-1102	Sequence 1102, Ap
C 17	20.8	43.3	1029	US-09-328-352-2205	Sequence 2205, Ap
C 18	20.6	42.9	1239	US-09-328-352-779	Sequence 779, App
C 19	20.6	42.9	1872	US-09-280-116-110	Sequence 110, App
C 20	20.6	42.9	3366	US-09-345-650-2	Sequence 2, Appl
C 21	20.6	42.9	8334	US-09-543-681A-1952	Sequence 1952, Ap
C 22	20.4	42.5	440	US-09-621-976-8341	Sequence 8341, Ap
C 23	20.4	42.5	753	US-09-252-991A-13564	Sequence 13564, A
C 24	20.4	42.5	765	US-09-252-991A-13860	Sequence 13860, A
C 25	20.4	42.5	1404	US-09-107-532A-2521	Sequence 2521, Ap
C 26	20.4	42.5	1608	US-09-328-352-2416	Sequence 2416, Ap
C 27	20.4	42.5	1848	US-08-637-670-23	Sequence 23, Appl

C 28	20.2	42.1	1203	US-09-107-532A-2612	Sequence 2612, Ap
C 29	20	41.7	562	US-09-328-111-468	Sequence 468, App
C 30	20	41.7	1152	US-09-489-039A-2755	Sequence 2755, Ap
C 31	20	41.7	1280	US-08-335-844A-10	Sequence 10, Appl
C 32	20	41.7	1280	US-09-129-366-10	Sequence 10, Appl
C 33	20	41.7	1293	US-08-335-844A-11	Sequence 11, Appl
C 34	20	41.7	1293	US-09-129-366-11	Sequence 11, Appl
C 35	20	41.7	2472	US-08-335-844A-7	Sequence 7, Appli
C 36	20	41.7	2472	US-09-129-366-7	Sequence 7, Appli
C 37	20	41.7	3213	US-09-107-532A-62	Sequence 62, Appl
C 38	20	41.7	3296	US-08-335-844A-14	Sequence 14, Appl
C 39	20	41.7	3296	US-09-129-366-14	Sequence 14, Appl
C 40	20	41.7	3305	US-08-335-844A-8	Sequence 8, Appli
C 41	20	41.7	3305	US-09-129-366-8	Sequence 8, Appli
C 42	20	41.7	3319	US-08-335-844A-15	Sequence 15, Appl
C 43	20	41.7	3319	US-09-129-366-15	Sequence 15, Appl
C 44	20	41.7	3358	US-08-335-844A-20	Sequence 20, Appl
C 45	20	41.7	3358	US-09-129-366-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1

US-08-961-527-78/C

Sequence 78, Application US/08961527

Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESSES:

ADDRESSES: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 3636 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-78

Query Match

Best Local Similarity 76.9%; Pred. No. 1.1;

Matches 30; Conservative 0; Mismatches 9; Indels 0;

6 TGAAGTGTACCGGTGTGTACCAATGAAATGCA 44

Db 2806 TCAACAGTGGCCCTGTGTGGCCCAATGATGCTCA 2768

RESULT 2

US-09-325-932A-15
Sequence 15, Application US/09325932A
Patent No. 6451604
GENERAL INFORMATION:
APPLICANT: Flinn, Barry
APPLICANT: Lasham, Annette
TITLE OF INVENTION: Compositions affecting programmed cell
FILE REFERENCE: 1022
CURRENT APPLICATION NUMBER: US/09/325,932A
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 206
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15
LENGTH: 1014
TYPE: DNA
ORGANISM: Pinus radiata
US-09-325-932A-15

Query Match

Best Local Similarity 46.7%; Score 22.4; DB 4; Length 1014;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 TATGATGAGCTGCTTACCGGTGTTATCCCATGAATGCGACATGCA 48
DB 439 TATGACGACGAGGCTTCTTCTTGACCTTGACCTGACATGCA 486

RESULT 3

US-08-961-527-166/c
Sequence 166, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae polynucleotides and sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 5406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-166

Query Match

45.8%; Score 22; DB 4; Length 5406;

Best Local Similarity 67.4%; Pred. No. 15;
Matches 31; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 ATGATCAGCTGGTTACCGGTGTTATCCCATGAATGCGACATGC 47
DB 3442 AAGATCAGCTAACACCGCTTATGATTACTCATCATATGGAATGC 3397

RESULT 4

US-09-634-238-3/c
Sequence 3, Application US/09634238
Patent No. 6544772
GENERAL INFORMATION:
APPLICANT: Glenn, Matthew
APPLICANT: Havukkala, Ilkka J.
APPLICANT: Bloksberg, Leonard, N.
APPLICANT: Lubbers, Mark W.
APPLICANT: Dekker, James
APPLICANT: Christenson, Anna C.
APPLICANT: Holland, Ross
APPLICANT: O'Toole, Paul W.
APPLICANT: Reid, Julian R.
TITLE OF INVENTION: Polynucleotides, materials incorporating
FILE REFERENCE: 11000,1043U1
CURRENT APPLICATION NUMBER: US/09/634,238
CURRENT FILING DATE: 2000-08-08
NUMBER OF SEQ ID NOS: 422
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 7754
TYPE: DNA
ORGANISM: Lactobacillus rhamnosus
US-09-634-238-3

Query Match

Best Local Similarity 45.4%; Score 21.8; DB 4; Length 7754;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 5 ATCAGCTGTTACCGGTGTTATCCCATGAATGCGACAT 45
DB 4023 AGCAGCGGCTTGTCTTACCGGTATTGGCATGAGACTGCCCAT 3983

RESULT 5

US-09-252-991A-11586/c
Sequence 11586, Application US/0925291A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 11586
LENGTH: 648
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11586

Query Match

Best Local Similarity 44.6%; Score 21.4; DB 4; Length 648;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 4 GATCAGCTGTTACCGGTGTTATCCCATGAATGCA 42

Db 296 GATCGCTCGATCCTGTTGTTTCCCATGCACTGGAA 258

RESULT 6

US-09-252-991A-11645/C
; Sequence 11645, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11645
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11645

Query Match 44.6%; Score 21.4; DB 4; Length 963;

Best Local Similarity 71.8%; Pred. No. 16; Mismatches 11; Indels 0; Gaps 0;

Db 316 GATCGCTCGATCCTGTTGTTTCCCATGCACTGGAA 278

Query 4 GATCGCTGTTACCGGTGTTTACCAGTAATGCA 42

US-09-252-991A-11794
; Sequence 11794, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11794
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11794

Query Match 44.6%; Score 21.4; DB 4; Length 987;

Best Local Similarity 71.8%; Pred. No. 16; Mismatches 11; Indels 0; Gaps 0;

Db 681 GATCGCTCGATCCTGTTGTTTCCCATGCACTGGAA 719

Query 4 GATCGCTGTTACCGGTGTTTACCAGTAATGCA 42

US-09-453-702B-36/C
; Sequence 36, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatterer, Frederick R.
; Burland, Valerie
; Perna, Nicole T.

Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESS: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 1983
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

US-09-453-702B-36
; Sequence 36, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatterer, Frederick R.
; Burland, Valerie
; Perna, Nicole T.

Query Match 44.6%; Score 21.4; DB 4; Length 1983;

Best Local Similarity 66.0%; Pred. No. 20; Mismatches 16; Indels 0; Gaps 0;

Db 372 ATGATCAATGACCACTTGTGTTACCGGTGTAATGCAATGCA 326

Query 2 ATGATCAATGACCACTTGTGTTACCGGTGTAATGCAATGCA 48

US-08-317-401E-3
; Sequence 3, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann
; TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLES OF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
ADDRESS: No. 5922561o No. 5922561th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (BRO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,401E

FILING DATE: 03-October-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Harrington, James J.
 REGISTRATION NUMBER: 38,711
 REFERENCE/DOCKET NUMBER: 4248.000-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 867 0298
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2166 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Aspergillus niger
 US-08-317-401E-3

Query Match 44.6%; Score 21.4; DB 2; Length 2166;
 Best Local Similarity 80.6%; Pred. No. 20;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 4 GATAGCTGTTACCCGTTGTTACCCATG 34
 Db 326 GATGAGCTGTTCCCTGTTGATCCCATG 356

RESULT 10
 US-08-317-401E-1
 Sequence 1, Application US/08317401E
 Patent No. 5922561
 GENERAL INFORMATION:
 APPLICANT: Thompson, Sheryl Ann
 APPLICANT: Javer, Debbie Sue
 TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 59225610 No. 5922561disk of No. 5922561th America, Inc.
 STREET: 405 Lexington Avenue, Suite 6400
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10174-6401
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25 (PPO)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/317,401E
 FILING DATE: 03-October-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Harrington, James J.
 REGISTRATION NUMBER: 38,711
 REFERENCE/DOCKET NUMBER: 4248.000-US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212 867 0123
 TELEFAX: 212 867 0298
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2877 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ORIGINAL SOURCE:
 ORGANISM: Aspergillus niger
 FEATURE:
 NAME/KEY: CDS
 LOCATION: join (126..203, 253..1776)
 US-08-317-401E-1

Query Match 44.6%; Score 21.4; DB 2; Length 2877;
 Best Local Similarity 80.6%; Pred. No. 22;
 Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
 Oy 4 GATAGCTGTTACCCGTTGTTACCCATG 34
 Db 418 GATGAGCTGTTCCCTGTTGATCCCATG 448

RESULT 11
 US-08-961-527-132/c
 Sequence 132, Application US/08961527
 Patent No. 6420135
 GENERAL INFORMATION:
 APPLICANT: Charles Kunsch
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 391
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,527
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 132:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 9541 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-527-132

Query Match 44.6%; Score 21.4; DB 4; Length 9541;
 Best Local Similarity 71.8%; Pred. No. 31;
 Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
 Oy 7 CAGCTGTTACCCGTTGTTACCCATGCAAT 45
 Db 609 CAGTTGATTAGCCGTTGTTACCCATGCAAT 6061

RESULT 12
 US-09-543-681A-3167
 Sequence 3167, Application US/09543681A
 Patent No. 6605709
 GENERAL INFORMATION:
 APPLICANT: GARY BRETON
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
 FILE REFERENCE: 2709.1002-001
 CURRENT APPLICATION NUMBER: US/09/543,681A
 PRIOR FILING DATE: 2000-04-05
 PRIOR APPLICATION NUMBER: US 60/128,706

PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 3167
LENGTH: 597
TYPE: DNA
ORGANISM: Proteus mirabilis
US-09-543-681A-3167

Query Match 43.8%; Score 21; DB 4; Length 597;
Best Local Similarity 73.0%; Pred. No. 21;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 12 GGTACCCGTTGTTACCATGAATGCACATGCA 48
Db 375 GGTGAACTGTGCTCACTGAAATGACACAGCA 411

RESULT 13
US-08-879-337-10
Sequence 10, Application US/08879337B
Patent No. 6639130
GENERAL INFORMATION:
APPLICANT: Jang, Yvan-Chyun
TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
FILE REFERENCE: 00786/338001
CURRENT APPLICATION NUMBER: US/08/879,337B
CURRENT FILING DATE: 1997-06-20
EARLIER APPLICATION NUMBER: 60/022,086
EARLIER FILING DATE: 1996-06-21
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10
LENGTH: 2975
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: variation
LOCATION: (1)...(2975)
OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-10

Query Match 43.8%; Score 21; DB 4; Length 2975;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTTACCCGTTGTTACCCATGAATGCACAT 45
Db 310 TTTATGAGATGGCTCTCACTTGTTCATGCGATGATATAT 354

RESULT 14
US-08-879-337-3
Sequence 3, Application US/08879337B
Patent No. 6639130
GENERAL INFORMATION:
APPLICANT: Jang, Yvan-Chyun
TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
FILE REFERENCE: 00786/338001
CURRENT APPLICATION NUMBER: US/08/879,337B
CURRENT FILING DATE: 1997-06-20
EARLIER APPLICATION NUMBER: 60/022,086
EARLIER FILING DATE: 1996-06-21
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 6587
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
NAME/KEY: variation
LOCATION: (1)...(6587)

OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-3

Query Match 43.8%; Score 21; DB 4; Length 6587;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTTACCCGTTGTTACCCATGAATGCACAT 45
Db 3664 TTTATGAGATGGCTCTCACTTGTTCATGCGATGATATAT 3708

RESULT 15
US-09-596-002-39
Sequence 39, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Patterson, Chandra
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
FILE REFERENCE: PM-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
CURRENT FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 39
LENGTH: 100848
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte template ID No. 6632636 39
PUBLICATION INFORMATION:
US-09-596-002-39

Query Match 43.8%; Score 21; DB 4; Length 100848;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 4 GATCAGCTGGTTACCCGTTGTTACCCATGAATGCACATGCA 48
Db 73642 GATGTGATTTGACACATATTTATCATCAATGCGACAGAA 73686

Search completed: March 16, 2004, 04:37:03
Job time : 7.01734 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 42.2775 seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48

Sequence: 1 taatgacagctgctacccg.....ccatgaatgcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCRT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCRTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27.8	57.9	2547	12	US-10-282-122A-36154 Sequence 36154, A
2	24.6	51.3	2547	12	US-10-282-122A-37710 Sequence 37710, A
3	24.6	51.3	3636	12	US-10-158-844-78 Sequence 78, App1
4	24	50.0	409	12	US-10-424-599-119735 Sequence 119735, A
5	23.2	48.3	720	12	US-10-425-114-6010 Sequence 6010, App
6	23	47.9	420	15	US-10-398-221-318 Sequence 318, App
7	23	47.9	420	15	US-10-398-221-3306 Sequence 2306, App
8	23	47.9	436	10	US-09-918-995-26274 Sequence 26274, A
9	23	47.9	3013020	15	US-10-398-221-1058 Sequence 10, App1
10	23	47.9	11677	12	US-10-282-122A-25322 Sequence 25322, App
11	22.6	47.1	1767	12	US-10-282-122A-25322 Sequence 25322, A
12	22.6	47.1	3984	15	US-10-369-493-42554 Sequence 42554, A
13	22.4	46.7	1014	14	US-10-219-220-15 Sequence 15, App1
14	22.4	46.7	1925	14	US-10-219-220-217 Sequence 217, App
15	22.2	46.2	912	9	US-09-738-626-2582 Sequence 2582, App

16	22.2	46.2	2370	15	US-10-369-493-32121 Sequence 32121, A
17	22.2	46.2	2487	15	US-10-369-493-44598 Sequence 44598, A
18	22.2	46.2	2574	15	US-10-369-493-34285 Sequence 34285, A
19	22.2	46.2	113604	14	US-10-227-195A-1 Sequence 1, App1
20	22.2	46.2	113604	14	US-10-227-195A-2 Sequence 2, App1
21	22.2	46.2	3309400	9	US-09-738-626-1 Sequence 1, App1
22	22.2	45.8	145	9	US-09-864-761-24629 Sequence 24629, A
23	22	45.8	480	14	US-10-326-956-320 Sequence 320, App
24	22	45.8	557	9	US-08-864-761-7925 Sequence 7925, App
25	22	45.8	1791	9	US-09-822-830A-241 Sequence 241, App
26	22	45.8	5406	12	US-10-158-844-166 Sequence 166, App
27	21.8	45.4	1470	12	US-10-424-599-56399 Sequence 56399, A
28	21.8	45.4	1510	12	US-10-425-114-36154 Sequence 36154, A
29	21.8	45.4	1646	12	US-10-425-114-31422 Sequence 31422, A
30	21.8	45.4	2394	12	US-10-425-114-32979 Sequence 32979, A
31	21.8	45.4	3328	12	US-10-425-114-34623 Sequence 34623, A
32	21.8	45.4	7754	15	US-10-264-213-3 Sequence 3, App1
33	21.8	45.4	8495	13	US-10-108-605-356 Sequence 356, App
34	21.6	45.0	633	9	US-09-917-800A-1050 Sequence 1050, App
35	21.6	45.0	729	9	US-09-738-626-1352 Sequence 1352, App
36	21.6	45.0	852	12	US-10-627-476-367 Sequence 367, App
37	21.6	45.0	1580	12	US-10-425-114-21821 Sequence 21821, A
38	21.6	45.0	1746	15	US-10-260-238-578 Sequence 578, App
39	21.6	45.0	2853	12	US-10-425-114-13682 Sequence 13682, A
40	21.6	45.0	1691139	14	US-10-067-514-1 Sequence 1, App1
41	21.6	45.0	1591139	15	US-10-419-723-1 Sequence 1, App1
42	21.6	45.0	3309400	9	US-09-738-626-1 Sequence 1, App1
43	21.4	44.6	1011	12	US-10-282-122A-30293 Sequence 30293, A
44	21.4	44.6	1179	10	US-09-769-744A-119 Sequence 119, App
45	21.4	44.6	1983	14	US-10-114-170-36 Sequence 36, App1

ALIGNMENTS

RESULT 1
US-10-282-122A-36154
Sequence 36154, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zykkind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EITRA.034A
CURRENT APPLICATION NUMBER: US/10/282, 122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191, 078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206, 848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207, 727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230, 335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230, 347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242, 578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253, 625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257, 931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267, 636

;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 36154
;; LENGTH: 2547
;; TYPE: DNA
;; ORGANISM: Streptococcus mutans
US-10-282-122A-36154

Query Match
Best Local Similarity 57.9%; Score 27.8; DB 12; Length 2547;
Matches 32; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 CAGCTGCTTACCGGTGTTTACCATGAATGGCAGCAT 45
DB 850 CAGCAGGTTCACCTGTGTGGCCCATGAACTGGCGCAT 888

RESULT 2
US-10-282-122A-37710
; Sequence 37710, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELPTRA.0344
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-32
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 37710
; LENGTH: 2547
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-37710

Query Match
Best Local Similarity 51.3%; Score 24.6; DB 12; Length 2547;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 6 TCAGCTGTTACCGGTGTTTACCATGAATGGCACA 44
DB 849 TCACAGTTCCTGTTGTGGCCCATGAATGGCTCA 887

RESULT 3
US-10-158-844-78/c
; Sequence 78, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-158-844-78

Query Match
Best Local Similarity 51.3%; Score 24.6; DB 12; Length 3636;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 6 TCAGCTGTTACCGGTGTTTACCATGAATGGCACA 44
DB 2806 TCACAGTTCCTGTTGTGGCCCATGAATGGCTCA 2768

RESULT 4
US-10-424-599-119735/c
; Sequence 119735, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: Ia Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 119735
LENGTH: 409
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_79129C.1
US-10-424-599-119735

Query Match
Best Local Similarity 50.0%; Score 24; DB 12; Length 409;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGACATGCA 48
Db 340 TACGGTACCACTTGCCTGTTCTCGCCGAGAAAGGCCATCCA 293

RESULT 5
US-10-425-114-6010
Sequence 6010, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jindong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 6010
LENGTH: 720
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700550972_FLI
US-10-425-114-6010

Query Match
Best Local Similarity 48.3%; Score 23.2; DB 12; Length 720;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGTTACCCGTTGTTTACCCATGAATGACATGCA 45
Db 285 ATGTCAGCTGTTCCGATCTTTTCCCATGAGAAAGCCCAT 328

RESULT 6
US-10-398-221-318
Sequence 318, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederick
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 318
LENGTH: 420
TYPE: DNA
ORGANISM: Listeria innocua
US-10-398-221-318

Query Match
Best Local Similarity 47.9%; Score 23; DB 15; Length 420;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGACATGCG 47
Db 175 TCTGATTATCAGTTAAGTATGTAGTCTCTCATGAACTAGGCGATGC 221

RESULT 7
US-10-398-221-2306
Sequence 2306, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederick
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2306
LENGTH: 420
TYPE: DNA
ORGANISM: Listeria innocua
US-10-398-221-2306

Query Match
Best Local Similarity 47.9%; Score 23; DB 15; Length 420;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGACATGCG 47
Db 175 TCTGATTATCAGTTAAGTATGTAGTCTCTCATGAACTAGGCGATGC 221

RESULT 8
US-09-918-995-26274/C
Sequence 26274, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 26274
LENGTH: 436
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)...(436)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26274

Query Match
Best Local Similarity 47.9%; Score 23; DB 10; Length 436;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGTTACCCGTTGTTTACCCATGAATGACATGCA 48
Db 175 TCTGATTATCAGTTAAGTATGTAGTCTCTCATGAACTAGGCGATGC 221

;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 42554
;; LENGTH: 3984
;; TYPE: DNA
;; ORGANISM: Anabaena PCC7120
US-10-369-493-42554

Query Match 47.1%; Score 22.6; DB 15; Length 3984;
Best Local Similarity 75.7%; Pred. No. 47;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 6 TCAGCTGGTACCCGCTGTGTTACCCATGAATGGCA 42
Db 843 TCACCTACTTACCGGTGATGATGCCATTATGTGCA 879

RESULT 13
US-10-219-220-15
; Sequence 15, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Laasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of plant development
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Pinus radiata
US-10-219-220-15

Query Match 46.7%; Score 22.4; DB 14; Length 1014;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTACCCGCTGTGTTACCCATGAATGGCAATGCA 48
Db 439 TATGACGAGGAGGCTTCTTGACAGTTTCACCTGAGACAGATGGA 486

RESULT 14
US-10-219-220-217
; Sequence 217, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Laasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of plant development
; FILE REFERENCE: 11000.1022c1
; CURRENT APPLICATION NUMBER: US/10/219,220
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
; PRIOR FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 290
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 217
; LENGTH: 1925
; TYPE: DNA
; ORGANISM: Pinus radiata
US-10-219-220-217

Query Match 46.7%; Score 22.4; DB 14; Length 1925;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTGTACCCGCTGTGTTACCCATGAATGGCAATGCA 48
Db 980 TATGACGAGGAGGCTTCTTGACAGTTTCACCTGAGACAGATGGA 1027

RESULT 15
US-09-738-626-2582
; Sequence 2582, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2582
; LENGTH: 912
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2582

Query Match 46.2%; Score 22.2; DB 9; Length 912;
Best Local Similarity 69.8%; Pred. No. 44;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 3 TGATCAGCTGTGTACCCGCTGTGTTACCCATGAATGGCAATGCAAT 45
Db 180 TGATGCGCTGTGTCAGCGTGTGTCGTCATGAGAGGACAGT 222

Search completed: March 17, 2004, 08:15:38
Job time : 54.3775 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 234.832 Seconds

(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48

Sequence: 1 tatgatcagctgcgttaccgc.....cccatgaatgcacatgca 48

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthm:*
3: em_estin:*
4: em_estmv:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc1:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc1:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rtd:*
26: em_ges_phg:*
27: em_ges_vrt1:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25.2	52.5	878	28	CC001571 PDIE23TD
2	25	52.1	894	13	BU147350 AGENCOURT
3	24.8	51.7	273	10	BB435111 BB435111
4	24.6	51.3	1201	9	AL576676 AL576676

5	24.2	50.4	446	9	AU299906
6	24.2	50.4	539	28	AZ056038
7	24.2	50.4	903	14	CB200568
8	24.2	50.4	1032	29	CNS01XV
9	24	50.0	176	10	AM837303
10	24	50.0	300	10	BF634308
11	24	50.0	600	12	BJ102563
12	24	50.0	610	9	AL589140
13	24	50.0	766	13	B0294276
14	24	50.0	780	13	B0252275
15	24	50.0	791	13	B0344056
16	24	50.0	812	13	B0328355
17	24	50.0	813	13	B0202575
18	23.8	49.6	294	9	AA224363
19	23.6	49.2	296	10	AW373707
20	23.6	49.2	316	14	CB017485
21	23.6	49.2	994	13	B0250654
22	23.4	48.8	446	10	BE443880
23	23.4	48.8	605	14	CB891544
24	23.4	48.8	672	14	CD452770
25	23.4	48.8	1016	29	CC600714
26	23.2	48.3	153	13	BQ172708
27	23.2	48.3	230	10	BB435292
28	23.2	48.3	248	14	CF245156
29	23.2	48.3	275	10	BB435052
30	23.2	48.3	295	10	BB436405
31	23.2	48.3	305	10	BB433896
32	23.2	48.3	460	12	BM379062
33	23.2	48.3	467	13	BQ744659
34	23.2	48.3	479	12	BG947566
35	23.2	48.3	521	14	CD204019
36	23.2	48.3	522	12	BI076348
37	23.2	48.3	540	9	A1711846
38	23.2	48.3	545	10	BE358333
39	23.2	48.3	577	14	CA404413
40	23.2	48.3	590	12	BG103593
41	23.2	48.3	591	12	BM318096
42	23.2	48.3	615	10	BE355335
43	23.2	48.3	620	14	CF049562
44	23.2	48.3	623	10	BE919209
45	23.2	48.3	623	14	CD527595

ALIGNMENTS

RESULT 1
CC001571
LOCUS
DEFINITION PDIE23TD_ZM_0.6_1.0_KB Zea mays genomic clone ZMMB217C21.
ACCESSION CC001571
VERSION CC001571.1 GI:29380131
KEYWORDS
SOURCE
ORGANISM Zea mays
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 878)
REFERENCE Whitehead, C.A., Quackenbush, J., Van Aken, S., Uterback, T.,
Reenick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Benneken, J.
Maize Genomics Consortium
Unpublished (2003)
CONTACT Contact: Cathy Whitelaw
TITLE
JOURNAL
COMMENT

712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitehead@tigr.org
Seq primer: TF
Class: sheared ends.

FEATURES

Location/Qualifiers

1. 878

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBTA217C21"

/clone_lib="ZM 0.6-1.0 kb"

/note="Vector: pCR4-TOPO, Site 1: EcoRI; 0.6-1.0 kb high"

Cor selected genomic DNA library"

ORIGIN

Query Match

Best Local Similarity 52.5%; Score 25.2; DB 28; Length 878;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

2 ATGATCAGCTGTTACCCGTTGTTACCAATGAAATGCAATGCA 47
548 ATGACCGACAGCTGTTCCACTGTAACCAATGAAATGCAATGCAATGCA 593

Db

RESULT 2

BUI47350

LOCUS

DEFINITION

AGENCOURT 8070809 NIH_MGC_112 Homo sapiens CDNA clone IMAGE:6089930

ACCESSION

BUI47350

VERSION

BUI47350.1 GI:22660882

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE

1 (bases 1 to 894) NIH-MGC http://mgs.nci.nih.gov/.

AUTHORS

Unpublished (1999)

TITLE

Contact: Robert Strausberg, Ph.D.

JOURNAL

Email: cgapds-remail.nih.gov

COMMENT

Tissue Procurement: DCTP/DRP

FEATURES

CDNA Library Preparation: Rubin Laboratory

source

DNA Sequencing by: Agencourt Bioscience Corporation

Location/Qualifiers

1. 894

organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6089930"

/lab host="DH10B (phage-resistant)"

/clone_lib="NIH MGC 112"

/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; CDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGCGAG(G). Library constructed by Ling Hong in the

laboratory of Gerald M. Rubin (University of California,

Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and

Superscript II RT (Life Technologies). Note: this is a

NIH_MGC library."

ORIGIN

Query Match

Best Local Similarity 52.1%; Score 25; DB 13; Length 894;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 ATGATAGCTGTTACCCGTTGTTACCAATGAAATGCA 42

Db

841 ATGACGACCCCTTACCACTGTTTACCAACCAAGGCGCA 881

RESULT 3

BB435111

LOCUS

BB435111

DEFINITION

BB435111 RIKEN full-length enriched, adult pancreas islet cells Mus

ACCESSION

BB435111

VERSION

BB435111.1 GI:9274838

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 273) Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,

AUTHORS

Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,

LOCUS

Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,

DEFINITION

Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K.,

ACCESSION

Shibata, K., Shibata, Y., Shigemoto, Y., Shingawa, A., Shiraki, T.,

VERSION

Sogabe, Y., Sugahara, Y., Suzuki, H., Tagawa, A., Tachibana, F.,

KEYWORDS

Takahashi, F., Tomimaga, N., Toyota, T., Tsunoda, Y., Watanabe, S.,

SOURCE

Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Yamashita, M., and

ORGANISM

RIKEN Mouse ESTs (Kono, H., et al.)

COMMENT

Unpublished (2000)

FEATURES

Contact: Yoshihide Hayashizaki

source

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Location/Qualifiers

The Institute of Physical and Chemical Research (RIKEN)

organism="Mus musculus"

1. 273

/db_xref="taxon:10090"

/clone="C820011N02"

/tissue_type="pancreas"

/cell_type="islet cells"

/dev stage="adult"

/lab host="DH10B"

/clone_lib="RIKEN full-length enriched, adult pancreas

islet cells"

/note="Site 1: SalI; Site 2: BamHI; CDNA library was

prepared and sequenced in Mouse Genome Encyclopedia

Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in

RIKEN. Division of Experimental Animal Research in Riken

contributed to prepare mouse tissues. 1st strand cDNA was

further details.

Please visit our web site (<http://genome.rtc.riken.go.jp>) for

Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCT-23. For BAC library availability, please contact Pieter de Jong (pieter@redjond.med.buffalo.edu). Clones may be purchased from BACPac Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tcd/bac/bac_ends/mouse/bac_end_intro.html
Plate: 413 row: M column: 16
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..539

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCT-23-413M16"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1: EcoRI, Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI MethyIase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match
Best Local Similarity 50.4%; Score 24.2; DB 28; Length 539;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GATCAGCTGTTACCGCTGTTGATCCATGAATGCGACATG 48
DB 493 GTTCAGATGTTGCAATGTTGACCCATTCATGTCACATGCA 449

RESULT 7
CB200568
LOCUS
DEFINITION
ABENCCURT 11271837 NICHDP_XCG_Tad2 xenopus laevis cDNA clone
IMAGE:6875681 5', mRNA sequence.
VERSION
CB200568
KEYWORDS
EST
CB200568.2 GI:29147154
SOURCE
Xenopus laevis (African clawed frog)
Xenopus laevis
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 903)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Feb 4, 2003 this sequence version replaced gi:28233219.
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Drs. Donald Brown and Liqun Cai
CDNA Library Preparation: CLONTECH
DNA Sequencing by: The I.M.A.G.E. Consortium (LIML)
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LHC311 row: h column: 16
High quality sequence stop: 104.
Location/Qualifiers
1..903
/organism="Xenopus laevis"
/mol_type="mRNA"

FEATURES
source

FEATURES

Location/Qualifiers

1..903

/organism="Xenopus laevis"

/mol_type="mRNA"

ORIGIN

Query Match
Best Local Similarity 50.4%; Score 24.2; DB 14; Length 903;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 ATGATCAGCTGTTACCGCTGTTGATCCATGAATGCGACATG 46
DB 434 ATGATGAGTCCCTACTCTATTGCTCAGTGAAGTCTACATG 478

RESULT 8
CNS01XV/C
LOCUS
DEFINITION
CNS01XV 1032 bp DNA linear GSS 01-SEP-2000
Tetradon nigroviridis genome survey sequence PUC-ori end of clone
217C04 of library G from Tetradon nigroviridis, genomic survey
sequence.
AL172228
AL172228.1 GI:7810285
GSS: genome survey sequence.
KEYWORDS
Tetradon nigroviridis
Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.
1
Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Mincker P., Brotier P., Quetier F.,
Saurin W. and Weissenbach J.
Estimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20296633
10835645

Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Mincker P., Brotier P., Quetier F.,
Saurin W., Bernot A., and Weissenbach J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

JOURNAL
MEDLINE
PUBMED
20296633
10835645

Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Mincker P., Brotier P., Quetier F.,
Saurin W., Bernot A., and Weissenbach J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

JOURNAL
MEDLINE
PUBMED
20296633
10835645

Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Mincker P., Brotier P., Quetier F.,
Saurin W., Bernot A., and Weissenbach J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

JOURNAL
MEDLINE
PUBMED
20296633
10835645

Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Mincker P., Brotier P., Quetier F.,
Saurin W., Bernot A., and Weissenbach J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Mincker P., Brotier P., Quetier F.,
Saurin W., Bernot A., and Weissenbach J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Mincker P., Brotier P., Quetier F.,
Saurin W., Bernot A., and Weissenbach J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

clone.lib="G"
/note="Genoscope sequence ID : C0AG217B02SP1-end :
PUC-Orl"

ORIGIN

Query Match 50.4%; Score 24.2; DB 29; Length 1032;
Best Local Similarity 71.1%; Pred. No. 2.2e+02;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 2 ATGATCAGCTGGTTACCCGTTGTTACCATGAATGCGACATG 46
Db 573 AAGCAGACTGGTTATCAGATTGTACACAGTGAAGAACTG 529

RESULT 9
AM837303 176 bp mRNA linear EST 18-MAY-2000
LOCUS QV2-LT0038-180300-079-d01 LT0038 Homo sapiens cDNA, mRNA sequence.
DEFINITION AM837303
ACCESSION AM837303
VERSION AM837303.1 GI:7931277
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 176)
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
Goldman, G.H., Carvalho, A.F., Matukuma, A., Bala, G.S., Simpson, D.H.,
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V.,
O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
Simpson, A.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the PAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=et2-QV2-LT0038-180
300-079-d01&t3=2000-03-18&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 26
High quality sequence stop: 176.
Location/Qualifiers
1..176
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="LT0038"
/note="Organ: leiomyos; Vector: puc18; Site 1: Sma1;
Site 2: Sma1; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

FEATURES
source

ORIGIN

Query Match 50.0%; Score 24; DB 10; Length 176;
Best Local Similarity 68.8%; Pred. No. 1.4e+02;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY 1 TATGATGAGTGCTTACCCGTTGTTATCCATGAATGACATGCA 48

Db

RESULT 10
BF634308 300 bp mRNA linear EST 19-DEC-2000
LOCUS NF082C02DP1F1017 Drought Medicago truncatula cDNA clone NF082C02DT
DEFINITION NF082C02DP1F1017 Drought Medicago truncatula cDNA clone NF082C02DT
5', mRNA sequence.
ACCESSION BF634308
VERSION BF634308.1 GI:11898466
KEYWORDS EST.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosoid I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.

REFERENCE 1 (bases 1 to 300)
Torrez-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,
Flores, H.R., Imman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
Unpublished (2000)

JOURNAL Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert Length: 300 Std Error: 0.00
Plate: 082 Row: C Column: 02
Seq primer: TCACACAGAAACAGCTATGAC.
Location/Qualifiers
1..300
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF082C02DT"
/tissue_type="plantlets"
/dev_stage="pooled timepoints"
/clone_lib="Drought"
/note="Vector: Lambda Zap; Contains a mixture of entire
plantlets harvested in a series of days-post-watering
timepoints."

FEATURES
source

ORIGIN

Query Match 50.0%; Score 24; DB 10; Length 300;
Best Local Similarity 73.2%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 5 ATCAGCTGTTACCCGTTGTTTACCATGAATGACAT 45
Db 221 AAGCAGAGGTTGCTGTTGTTATCCATGACATGACAT 261

RESULT 11
B1102563 600 bp mRNA linear EST 30-MAY-2003
LOCUS B1102563
DEFINITION B1102563 unpublished oligo-capped cDNA library, C. elegans L1 stage
Caenorhabditis elegans cDNA clone Yk1040a11 5', mRNA sequence.
ACCESSION B1102563
VERSION B1102563.2 GI:31238878
KEYWORDS EST.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Pelodexinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 600)
Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
and Sugano, S.
A complementary view of the C.elegans genome

TITLE
REFERENCE
AUTHORS
A complementary view of the C.elegans genome

THIS PAGE BLANK (USPTO)

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

•

1
2
3
4
5
6
7
8
9
10
11
12

U43280 Salm

AE008749 Sa

AE016840 Sa

I44908 Sequ

AJ515701 Ci

X90754 E.CO

D90741 Esch

L04979 EBch

AF275733 B8

AP002554 ES

AE015131 Sh

AF237726 Sh

D90742 Esch

AX655393 Se

AF343445 La

AC146009 Pa

AE017038 Ba

ACI46008 Pa

AC009716 HO

AF325159 P1
AF325160 P1

AF325161 P1

AC092327 Ho
AY059943 Oy

PAT 07

Bloumher, S.C.

	Matches	424;	Conservative	0;	Mismatches	32;	Indels	0;	Gaps	0;
OY	1	ATGAAACCTTTTAAAGGTGGCAGCATTCGCAGCAATCGTAGTTCTTGCGAGTGTCTGCT	60							
Dd	1193	ATGAAACTTTTTAAAAGTGGCAGCATTTCCACGCAATCGTAGTTCTTGCGAGTGTCTGCT	1252							
OY	61	GCGCGCTTCCACAAATGGGGCGGCGCGGTAACTCAATACGGCGGCGGCAATAGTTCGGGC	120							
Dd	1253	GCGCGCTTCCACAATGGGGCGGCGCGGTAACTCAATACGGCGGCGGCAATAGTTCGGGC	1312							
OY	121	CCGGACTCAAAGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTGTCTGCGAA	180							
Dd	1313	CCGGACTCAAAGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTGTCTGCGAA	1372							
OY	181	AGCGATGCCCGTAATCTGAAACGACCAATTACCAGACGGGTATGTTAAGGCGCCGAT	240							
Dd	1373	AGCGATGCCCGTAATCTGAAACGACCAATTACCAGACGGGTATGTTAAGGCGCCGAT	1433							
OY	241	GTAAGCCAGGGTCCGATPATATGTAATACTGACTCAGAATGTTTCAGAAATAT	300							
Dd	1433	GTAAGCCAGGGTCCGATPATATGTAATACTGACTCAGAATGTTTCAGAAATAT	1492							
OY	301	GCCACCAATCGACCAAGTGGACGCTAAATCCGATATTAATCTGTCGGCCAATACGGCGGT	360							
Dd	1493	GCCACCAATCGACCAAGTGGACGCTAAATCCGATATTAATCTGTCGGCCAATACGGCGGT	1552							
OY	361	AATAACGCGCGCTGTGTTAATTATGATCAGCTGTTAACCCGTGTGTTAATCCAGTAATG	420							
Dd	1553	AATAACGCGCGCTGTGTTAATCAGACCGCAATCTGATTCAGAGTTAATGTGTGATGTT	1612							
OY	421	GCACATGCAACAAACGCCAGGCTAACCAAGATTAA	456							
Dd	1613	GCTTTTGCAACAAACGCCAGGCTAACCAAGATTAA	1648							

```

RESULT 3
STAJ2301
LOCUS          STAJ2301          5103 bp    DNA       linear   BCT 15-NOV-2000
DEFINITION    Salmomella typhimurium cs9c, cs9f, cs9E, cs9D, cs9B, cs9A, and 6
              cs9c genes.
ACCESSION     AJ002301
VERSION       AJ002301.1  GI:2739232
KEYWORDS      cs9a gene; cs9B gene; cs9c gene; cs9d gene; cs9f gene;
              cs9g gene.
SOURCE        Salmomella typhimurium
              Salmomella typhimurium
              Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
              Enterobacteriaceae; Salmomella.
REFERENCE     1
AUTHORS       Romling, U., Bian, Z., Hammar, M., Sterralta, W.D. and Normark, S.
TITLE         Culi1 fibers are highly conserved between Salmomella typhimurium
              and Escherichia coli with respect to operon structure and
              regulation
JOURNAL       J. Bacteriol. 180 (3), 722-731 (1998)
MEDLINE       98117058
PUBMED        9457880
REFERENCE     2 (bases 1 to 5103)
AUTHORS       Romling, U.
TITLE         Direct Submission
JOURNAL       Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
              Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES     location/Qualifiers
              1..5103
               /organism="Salmomella typhimurium"
               /mol_type="genomic DNA"
               /strain="SR-11"
               /db_xref="taxon:602"
               <1..322
               /note="ORF179a"
               /codon_start=2
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="CA05311.1"

```

gene
complement(361. .1194)
/gene="csgg"
complement(361. .1194)
/gene="csgg"
/function="lipoprotein"
/codon_start=1
evidence=experimental
/transl_table=11
/product="CsgG protein"
/protein_id="CA05312.1"
/db_xref="GI:2739234"
/db_xref="GOA:O54291"
/db_xref="SWISS-PROT:O54291"
/translation="MPRLILVAVLISGLCTAPPKQAAKPTLMPRAOSYKDIITLHPA
PGKIFVSVYNIDETGFQFKFYPASNESTAVPOSATMLVTALKDSMFPLEOGIQL
NLNRKRILIRPAENGTAAMNRIPLQSILFAANWGSIIIGESNVKSQGVGARVYG
IGADYOLODIAVNLRVNVSGBEILSSVTSITLIISYVOAGVPEDIOYLLBEE
IGTISENPMLCLMSAITGYIFLINSDIGRGLMDLNKKDRNDILVKTRELSVPE
S"

CDS
complement(1221. .1637)
/gene="csgf"
complement(1221. .1637)
/gene="csgf"
/codon_start=1
/transl_table=11
/product="CsgF protein"
/protein_id="CA05313.1"
/db_xref="GI:2739235"
/db_xref="SWISS-PROT:O54292"
/translation="MRVKHAYVLLMLRSPLTWAGNTFOEPNNPFGGNPNNGSLINS
AOANSYDDPYANDFGIEFPSALDNFTQAIOQILOGLLTNINTGRGRMVTDFFLI
DLNRDQGLQJLVNVDRTKGTSTIEVSGLTQTSTD"

gene
complement(1664. .2059)
/gene="csgx"
complement(1664. .2059)
/gene="csgx"
/codon_start=1
/transl_table=11
/product="CsgX protein"
/protein_id="CA05314.1"
/db_xref="GI:2739236"
/db_xref="SWISS-PROT:O54293"
/translation="MKRYLTWIVAELLFATGNIHANEEVEVPGILLTDHTVSSIGH
FRAPSDMSESYGNLTINERPARMSGMITTVNDVIFQFLPFMRKDPEXTVVF
ALAQTEALNRROLDQTLTSLRDARBF"

CDS
complement(2064. .2714)
/gene="csgd"
complement(2064. .2714)
/gene="csgd"
/function="transcriptional regulator"
/codon_start=1
/transl_table=11
/product="CsgD protein"
/protein_id="CA05315.1"
/db_xref="GI:2739237"
/db_xref="GOA:O54294"
/db_xref="SWISS-PROT:O54294"
/translation="MENEVSHGHITLLITKESLAQTALLQHKOSLATGGLAHNQ
RSLEITSAGCVILMDNMEADKELIHYYQDNLSRRNNKITULNTLPDPIYREIEWP
HINGPVATEQDEHVSGLOGLINGECYFQKLASYLTISGVRYNSTESALTIRE
KEILNKAKIRIGASNEIARSLEPSENTVTKTHLYNLFKIAVNKETOAVSAMNDLR"

gene
complement(2869. .2896)
/gene="csgDFG"
complement(2869. .2874)
/gene="csgDFG"
complement(2891. .2896)
/gene="csgDFG"

-10_signal
-35_signal

```

/translation="MSVKKKNIPTAIGCTICAFHSHAVGQGVVGGVTHFNGALVEPL
CDISTHAENI"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:O54295"
/protein_id="CAA05319.1"
/translation="MSVKKKNIPTAIGCTICAFHSHAVGQGVVGGVTHFNGALVEPL
CDISTHAENI"

```

Query Match	87.0%	Score 396.8	DB 1	Length 5103
Best Local Similarity	91.9%	Pred. No. 1,3e-101		
Matches 419	Conservative 0	Mismatches 37	Indels 0	Gaps 0
QY	1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTATGTTCTCGCAGTCTTGACT			60
Db	3966 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTATGTTCTCGCAGTCTTGACT			402
QY	61 GGGCGTGTTCACAATGGGGCGCGCGGTAAATCAATACGGCGCGCAATAGTTCGGC			120
Db	4026 GGGCGTGTTCACAATGGGGCGCGCGGTAAATCAATACGGCGCGCAATAGTTCGGC			408
QY	121 CCGGACTCAAGTTAGCAATTATCAGACGGTTCGGTAAACGTCGCTTCTGCAA			180
Db	4086 CCGGACTCAAGTTAGCAATTATCAGACGGTTCGGTAAACGTCGCTTCTGCAA			414
QY	181 ACGCATGCCCGTAAATCTGAAACGACATTAACCAAGCGGTATGTAAACGGCGCGAT			240
Db	4146 ACGCATGCCCGTAAATCTGAAACGACATTAACCAAGCGGTATGTAAACGGCGCGAT			420
QY	241 GTAGGCGCGGGTGGCAATTAAGTACTATTTGAAGTACTGACATGCGTTCGAAATAT			300
Db	4206 GTAGGCGCGGGTGGCAATTAAGTACTATTTGAAGTACTGACATGCGTTCGAAATAT			426
QY	301 GCCACCATCGACAGTGGGAACGCTAAATCTCGATATTAATCTGTCGGCAATACGGCGGT			360
Db	4286 GCCACCATCGACAGTGGGAACGCTAAATCTCGATATTAATCTGTCGGCAATACGGCGGT			4325
QY	361 AATTAACCGCGCGGTGTTAATTAATGATGACGCTGTACCGGTGTTATACCAATGAATG			420
Db	4326 AATTAACCGCGCGGTGTTAATGACCGCATCTGATTCACACGTAATGATGCGTCAATG			4385
QY	421 GCACATGCAACACGCCACGCGCTAACAGATTTAA			456
Db	4386 GGTATTGGCAACACGCCACGCGCTAACAGATTTAA			4421
RESULT 4	AE008749	22411 bp	DNA	Linear
LOCUS	Salmonella typhimurium LT2, section 53 of 220 of the complete			
ACCESSION	AE008749	AE006468		
KEYWORDS	AE008749.1	GI:16419641		
SOURCE	Salmonella typhimurium LT2			
ORGANISM	Salmonella typhimurium LT2			
REFERENCE	Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
AUTHORS	1 (bases 1 to 22411)			
TITLE	McClintell, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,			
JOURNAL	Letellier, P., Courtney, D., Portwiliak, S., Ali, O., Dante, M., Du, F.,			
MEDLINE	Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,			
PUBMED	Grewal, N., Mulvaney, E., Ryan, F., Sun, H., Flores, L., Miller, W.,			
REFERENCE	Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.			
AUTHORS	Complete genome sequence of Salmonella enterica serovar Typhimurium			
CONSTRM	LT2			
TITLE	Nature 413 (6658), 852-856 (2001)			
JOURNAL	Medline 11534948			
MEDLINE	11677609			
PUBMED	2 (bases 1 to 22411)			
REFERENCE	The Salmonella typhimurium Genome Sequencing Project			
AUTHORS	Direct Submission			
CONSTRM	Submitted (29-MAR-2001)			
TITLE				
JOURNAL				

CDS

6805..7497
/gene="STM1131"
/note="similar to E. coli orf. hypothetical protein
identity in aa 4 - 239"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein"
/protein_id="AA020062.1"
/db_xref="GI:16419647"
/translation="MKINKYLGVMSFIASSYLQATLIDYHEATDTIRNKORAI
IESKLPDPIGIVYDASVSGVDEODKHLSDIVANAELIGSYNKVNDVLPDGI
DELKLEYNFTWMSDQKFDNKNTVNEHVALAMKLNSTPYVEGVAVRNTDER
QTRRVGLQYH"
/gene="STM1132"
7768..7773
/gene="STM1132"
/note="putative RBS for STM1132; RegulonDB:STMS1H001403"
/db_xref="GI:16419647"
/translation=""
/gene="STM1132"
7780..9060
/note="similar to E. coli putative transport protein
identity in aa 25 - 418"
/codon_start=1
/transl_table=11
/product="putative sugar transport protein"

Query Match

Best Local Similarity 87.0%; Score 396.8; DB 1; Length 22411;
Matches 419; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY	1	ATGAACTTTAAAGTGGCAGCATTTGGCAGCAATGTMAGTTCTGGAGTCTGCTGCT	60
DB	17769	ATGAACTTTAAAGTGGCAGCATTTGGCAGCAATGTMAGTTCTGGAGTCTGCTGCT	60
QY	61	GGCGTCTTCCACAAATGGGGCGGGCGGCTTAATCAATAGCGGGCGCAATAGTTCCGGC	120
DB	17829	GGCGTCTTCCACAAATGGGGCGGGCGGCTTAATCAATAGCGGGCGCAATAGTTCCGGC	120
QY	121	CCGGACTCAAGTGTGAGCAATTAATGATGATGATGATGATGATGATGATGATGATGAT	180
DB	17889	CCGGACTCAAGTGTGAGCAATTAATGATGATGATGATGATGATGATGATGATGATGAT	180
QY	181	AGGATGCGCGTAAATCTGAAAAGCAATTAATGATGATGATGATGATGATGATGATGAT	240
DB	17949	AGGATGCGCGTAAATCTGAAAAGCAATTAATGATGATGATGATGATGATGATGATGAT	240
QY	241	GTAAGCGCAGGTCGGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT	300
DB	18009	GTAAGCGCAGGTCGGATTAATGATGATGATGATGATGATGATGATGATGATGATGAT	300
QY	301	GGCACCACGACGACGATGAAAGCTGCAATTAATGATGATGATGATGATGATGATGAT	360
DB	18069	GGCACCACGACGACGATGAAAGCTGCAATTAATGATGATGATGATGATGATGATGAT	360
QY	361	AATAACGGCGGCTGCTTAATGATGATGATGATGATGATGATGATGATGATGATGAT	420
DB	18129	AATAACGGCGGCTGCTTAATGATGATGATGATGATGATGATGATGATGATGATGAT	420
QY	421	GCAATGCAAAACAGCGGACGCTTAACGATTAATGATGATGATGATGATGATGATGAT	480
DB	18189	GCAATGCAAAACAGCGGACGCTTAACGATTAATGATGATGATGATGATGATGATGAT	480

RESULT 5

AL627269 254050 bp DNA linear BCT 04-JUN-2003
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome, segment 5/20.
ACCESSION AL627269 AL513382
VERSION AL627269.1 GI:16502231

KEYWORDS

Salmonella enterica subsp. enterica serovar Typhi

SOURCE

Salmonella enterica subsp. enterica serovar Typhi

ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

REFERENCE

1 (bases 1 to 254050)

AUTHORS

Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,

TITLE

Complete genome sequence of a multiple drug resistant Salmonella

JOURNAL

Nature 413 (6858), 848-852 (2001)

MEDLINE

21534947

PUBMED

11677608

REFERENCE

2 (bases 1 to 254050)

AUTHORS

Parkhill, J.

TITLE

Direct Submission

JOURNAL

Submitted (25-OCT-2001) Submitted on behalf of the Salmonella

COMMENT

sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

FEATURES

Details of S. Typhi sequencing at the Sanger Centre are available

source

Location/Qualifiers

FEATURES

1..254050

gene

/organism="Salmonella enterica subsp. enterica serovar

CDS

/note="Orthologue of E. coli pepN (AMPN ECOLI); Pasta hlt

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

gene /note="P500142 Neutral zinc metalloproteinases, zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q8Z759"
/db_xref="Swiss-Prot:Q8Z759"
/translation="MYPFVKALFQDLPERAHEFTFOQLRRITGTPLDALVRQKVP
KPTCMGLTFKPLGLAAGLDDECDALGAMGSLGIVTTPRPOGDKPLFR
LVDAEGLINRKNPNNLGVNLYENYKKAHPGIIIGINIKKNDTPVEGKDYDIME
KVYAVAGYIAINISPNRGLRTLOYGDALDLDLTJAIKKNKNDQVHHKXVPAVKI
APDCBEELIOVADSLRNHIDGVIATTTTDRSLVQMKCCQGTGLSGRPLOKST
EIIIRLSLELKGQPLIGVGIDSVIAREKIAAGATLVQISGPIFKGPLIKEIYV
HI"
misc_feature 3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODHase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
3238..3297
/gene="STY1079"
/note="P500911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="P500912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q8Z758"
/translation="MMRIKPDNNRWYDEEDHMMMLDLANGMLFRSRFSKMLTPPA
FCTPGFVDNALYFSPFEKCRDELTEKORALVNLALVIRYLKQMPKSMHFAVA
GEMWTPGDAASVMSLDTAEQVNLVYEPBENALCLIAOPGVVIAGRTNQLDAIK
IMNRLKPOVHCHSFSLEQAV"
complement(4716..5825)
/gene="STY1081"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q8Z757"
/db_xref="SPTREMBL:Q8Z757"
/translation="WATLSRLFIHPVKSRCIGLTHALADISGLAFRIEWTESDGT
FITARQPMWRFTPSPLDGLHFLTPRVSALVRFDTQDAPTEWGNHFAFARA
PTAINOMLSGFSSFDVOLCWGQPOLTRRVKHNVAVPGFAGGYLLTNESSLDOO
RCPAGVMEORPMLVYSGVAWEKDSKURLIGDVIEDVYKPCSRICFTTNSPEKQ
KHPSGELATIQARTADNNDVDFGQULIARNSCVIRVCEVITLATAKAYGATT
LDQSVTEPKHDSVITDMQGTFCGNNQVLLLEQLENGRIRIYSCAGIGCCRIR
LLEGEVPLKKSAGSDGTIISCSVPKTAURLN"
complement(4737..4955)
/gene="STY1081"
misc_feature

gene /note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein H10116 H10116 SW:YCBY_HAEIN (P43524; P43945) faeta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q8Z756"
/db_xref="SPTREMBL:Q8Z756"
/translation="MNSLPASTAGSLBKIELEKAGVCCVYGVGVRHRCPTRLI
YSLMMSRLASRIILPMGECVYSDDLIFGVQAINNTEIFNPATRAVHPSGLDITI
RNSQYGMKVKDALVDAFTKRLPRPVNDESPDLINVLNKTASIALDLSGDHL
LRGDRRTGLAPIKETLAAAIYMSGQPGTPLDPMCGSGTLLIEAMNATDRAPIGL
HRGHMGSQWAOHDEAIWQSKAEOARAKGLAEYSHSHPGSDSDARVIERASNNR
RAGIGELITFEVKVDAQVLSNPLPKPGVTVISNPVGBRVDSDPALIATSLGRTYK
NORGWMLSPFASPDLSGLQADQPAKKNPLDVCQNYHIAETADSKATYA
EDIANRRLKRLKLEKMARQEGIECTVLDADDEIVAAVDROYDAVIDAPKTV
DAQKARQRLFDIIAATLSVGIIPNNKVLITREKQKNQYOKMSKEGFLSEBYNA
RLWNLITDYLDITGLFDLRIARMLGMSKQPLNFSTYGSASVHAGLGAGANTTT
VDMSTRYLEWAERNLRLNGLSRAHRLIQDDCGLMLEANEOPLIFIDPPTFENSKR
MESEFDVDRDHVALLMDLRLKRLKRGGITMFSNNRGRMDLEGLAEGLTAQETIQT
LSDFPARNROIHNCWLIRAA"
misc_feature 5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="P501261 Uncharacterized protein family UPF0020 signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBT_ECOLI (530 aa), 32% identity in 526 aa overlap
Faeta hit to YJW_ECOLI (554 aa), 34% identity in 524 aa overlap
Query Match 86.7%; Score 395.2; DB 1; Length 254050;
Best Local Similarity 91.7%; Pred. No. 3.9e-101;
Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGNGGACGACATTCGACGACATCGATGTTTGCAGTCTGGCT 60
DB 88906 ATGAACCTTTAAAGNGGACGACATTCGACGACATCGATGTTTGCAGTCTGGCT 88965
QY 61 GGCCTCGTTCACAAATGGGGCGGGCGGTAAATCAATACGCGCGCAATAGTTCGGC 120
DB 88966 GGCCTCGTTCACAAATGGGGCGGGCGGTAAATCAATACGCGCGCAATAGTTCGGC 89025
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTGGCTTCTCTGCA 180
DB 89026 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTGGCTTCTCTGCA 89085
QY 181 AGCGATCCCGTAAATCGTGAAGACACATTACCGAGCGGTTATYGGTAACGGCGCGAT 240
DB 89086 AGCGATCCCGTAAATCGTGAAGACACATTACCGAGCGGTTATYGGTAACGGCGCGAT 89145
QY 241 GTAGCCAGGCGGTGGATTAATAGTACTATTGAAGTGAATCGAATGCTTTCAGAAATAT 300
DB 89146 GTAGCCAGGCGGTGGATTAATAGTACTATTGAAGTGAATCGAATGCTTTCAGAAATAT 89205

gene RAUNALGATPMDLMSILQSMOAGCLAKLEIT"
complement (5597..6295)
/gene="flgH"
/locus_tag="t1740"
complement (5597..6295)
/gene="flgH"
/locus_tag="t1740"
/notes="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MOKYLAHAPYVMAALMTLTGCAWIPAKPLVQAGTTAQPITPGV
PVANGSI.FQSAQPTINQYQPLFEDRPRLNIDTLIVLQENVASKSSAASADGKT
SFGEDIVRYLQGLFNGSRADMSGNSFGKGAASNTFSGTLITVTDQVLANN
LHVVGKQIAINQGTETFRFGSVNPRITISGNSVPSQVADARIEYVNGYINEAQN
MGMLQRFELNLSPM"
complement (6350..7132)
/gene="flgG"
/locus_tag="t1741"
complement (6350..7132)
/gene="flgG"
/locus_tag="t1741"
/notes="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flgG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISSLWIAKTGLDAQCTNMDVIANNLANSSTNGPKQRAVFEDL
LYQTRPGAQSSQITLPSGLQIGTVRPVATERLSQSLQSTNNSKQVAKKGQF
FOVWLPGTSAYRDSGFQVONQGLVAGFOQPAITIPANLSTIGSDVSVT
QOQCAAPVQGLMTTFMNDTLESIGENLYETQSSGAPNESTPLANGAGLLYQGY
VETSNVVAELVMIQVORAVEINSKAVSTTDQMLQKLTQL"
complement (7146..7901)
/gene="flgF"
/locus_tag="t1742"
complement (7146..7901)
/gene="flgF"
/locus_tag="t1742"
/notes="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="WDAIITYMGASQTLNQAVTNSNLANSSTPGFRAQLNLRAY
PVGLSLATRTLVASTPGADMTQGLDYTSRPLDALQDDGLVVOAADGAGYTEN
GNVOGPTQGLTIQGHVIGEGSPITVPESEITIAADGTISALNPEDPNTVAVER
LKLVAEGNEVQSDDLFRLLTAQAQERASVLAADPSIRIMSGVLEGSNVKPEAVMT
DMIANARFPMOKVITSVDNENGRANQLLSMS"
complement (7922..9133)
/gene="flgE"
/locus_tag="t1743"
complement (7922..9133)
/gene="flgE"
/locus_tag="t1743"
/notes="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFSQAVSGANAAATLVDVIGNNIANSATYRPSKTSAPADMA
GSRVGLGVKAVGITQDFTDGTNTTNGDLVAISQNGFRLVDSNGSVFYSRNGQFCL

Query Match 86.7%; Score 395.2; DB 1; Length 301983;
Best Local Similarity 91.7%; Pred. No. 3.9e-101;

Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACAAATCGTAGTTCTGCGAGTCTGCT 60
Db 37210 ATGAACTTTTAAAGTGGCAGCATTCGACAAATCGTAGTTCTGCGAGTCTGCT 37251
QY 61 GGCCTCGTTCACAAATGGGGGGGGGGGTAAATCAACGGCGCGCAATAGTCCGGC 120
Db 37250 GGCCTCGTTCACAAATGGGGGGGGGGGTAAATCAACGGCGCGCAATAGTCCGGC 37191
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCTCTGCA 180
Db 37190 CCGGATTCACGTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCTCTGCA 37131
QY 181 AGCGATCCCGTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
Db 37130 AGCGATCCCGTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 37071
QY 241 GTAGCCAGGGTGGGATTAATAGTACTTGAAGTGAAGTGGTTTCAGAAATAT 300
Db 37070 GTAGCCAGGGTGGGATTAATAGTACTTGAAGTGAAGTGGTTTCAGAAATAT 37011
QY 301 GCCACATCGACCGAGTGAAGCTAAATCCGATATTAAGTGGGCAATACGGCGGT 360
Db 37010 GCCACATCGACCGAGTGAAGCTAAATCCGATATTAAGTGGGCAATACGGCGGT 36951
QY 361 AATAAGCCGCGCTGTTAATATGATCAGCTGGTTACCGTGTGTTACCATGAATG 420
Db 36950 AATAAGCCGCGCTGTTAATATGATCAGCTGGTTACCGTGTGTTACCATGAATG 36891
QY 421 GCACATCAAAACGCGCAGCGCTAACCATTAAT 456
Db 36890 GCTTTTGCAACACGCGCAGCGCTAACCATTAAT 36855
RESULT 7
STAGEBA
LOCUS STAGEBA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AJ000514
VERSION AJ000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1
Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
Normark,S.J. and Rhen,M.
Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
Infect. Immun. 65 (12), 5320-5325 (1997)
JOURNAL
MEDLINE 98053981
PUBMED 9393832
REFERENCE 2 (bases 1 to 1048)
AUTHORS Sukupolvi,S.S.
TITLE Direct Submision
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinanmyllynkatu, 20520, FINLAND
location/Qualifiers
FEATURES
source
1..1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="AgfB protein"
/protein_id="CA04150.1"

```

/db_xref="GI:2275120"
/db_xref="SPTREMBL:O33801"
/translation="MKKLLFPMVLLTIGEPITATNTYDLARSEVNFPAVELSKSFNO
AAIGVGTDSARVRHSGSKLLSVISQERGNARRAVDQGNFAYIEQGNANDA
SISOSAYGNSAISAITOKSGSKANITQYQKTAVVVQKSHMAIQAINTQYGTG
KTAVVVQKSHMAIRVTR"
593..1048
/gene="agfa"
/cds
593..1048
/gene="agfa"
/codon_start=1
/codon_table=11
/product="Agfa protein"
/protein_id="CA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:O33802"
/translation="MKLLKVAFAALVSGSAVAVGVPMQGGGNNHNGGNSGPDST
LSIYQGSAAALALQSDARKSEHTITQSGYNGADVDGADNSTIELTQGFNNAT
IDQMNARSDITVGYGNNALVNTASDSVWVRQFNNAPANQYN"
ORIGIN

```

```

Query Match
Best Local Similarity 85.4%; Score 389.4; DB 1; Length 1048;
Matches 414; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```

```

QY 1 ATGAACCTTTAAAGTGGACGATTCACGCAATCGTAGTTCTGGCAGTCTGGCT 60
DB 593 ATGAACCTTTAAAGTGGACGATTCACGCAATCGTAGTTCTGGCAGTCTGGCT 60
QY 61 GGGCTGTCTCAAAATGGGGGGGGGGGCGGTAATATACGGCGGCAATAGTTCCGGC 652
DB 653 GGGCTGTCTCAAAATGGGGGGGGGGGCGGTAATATACGGCGGCAATAGTTCCGGC 652
QY 121 CCGGACTCAAGTGGACGATTCACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 120
DB 121 CCGGACTCAAGTGGACGATTCACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 120
QY 713 CCGGATTCAGTGGACGATTCACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 180
DB 713 CCGGATTCAGTGGACGATTCACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 180
QY 181 AGGATGCGCCCTAAATCTGAACGACCATTCACGAGGGGTATGGTAACGGCGCGAT 772
DB 773 AGGATGCGCCCTAAATCTGAACGACCATTCACGAGGGGTATGGTAACGGCGCGAT 772
QY 241 GTAGCGCAGGATCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 832
DB 241 GTAGCGCAGGATCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 832
QY 833 GTAGCGCAGGATCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
DB 833 GTAGCGCAGGATCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
QY 301 GCCACCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 892
DB 893 GCCACCATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 892
QY 361 AATAACGCGCGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 952
DB 953 AATAACGCGCGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 952
QY 421 GCACATGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 455
DB 421 GCACATGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 455
QY 1013 GGTTCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1047
DB 1013 GGTTCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1047

```

```

RESULT 8
LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 56 from patent US 5635617.
ACCESSION 144908
VERSION 144908.1 GI:2469621
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 361)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
JOURNAL Salmonella
Patent: US 5635617-A 56 03-JUN-1997;

```

```

FEATURES
Source Location/Qualifiers
1..361
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

```

```

Query Match
Best Local Similarity 69.1%; Score 315.2; DB 6; Length 361;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 64 GTGCTTCCAAATGGGGGGGGGGGCGGTAATATCAATACGGCGGCAATAGTTCCGGCCG 123
DB 1 GTGCTTCCAAATGGGGGGGGGGGCGGTAATATCAATACGGCGGCAATAGTTCCGGCCG 123
QY 124 GATCCACGTTGACGATTTATCAATACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 183
DB 61 GATCCACGTTGACGATTTATCAATACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 183
QY 184 GATCCACGTTGACGATTTATCAATACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 120
DB 121 GATCCACGTTGACGATTTATCAATACGATTCAGTTCGGCTAACGCTGCGCTCTCTGGAA 120
QY 244 GGGCAGGCGCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 303
DB 181 GGGCAGGCGCGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 303
QY 304 ACCATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 240
DB 241 ACCATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 240
QY 364 AACGCGCGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 383
DB 301 AACGCGCGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 383

```

```

RESULT 9
LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter sp. Fec2 csbA gene, csbA gene and csbA gene.
ACCESSION AU515700
VERSION AU515700.1 GI:31790491
KEYWORDS csbA gene; csbA gene; csbA gene; curlin-csbA protein; nucleation
component of curlin monomers; regulatory protein.
SOURCE
ORGANISM Enterobacteriaceae; Citrobacter.
REFERENCE 1
AUTHORS Zogaj,X., Bokranz,W., Nitz,M., and Romling,U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family
JOURNAL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
REFERENCE 2 (bases 1 to 2889)
AUTHORS Romling,U.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
Source Location/Qualifiers
1..2889
/organism="Citrobacter sp. Fec2"
/mol_type="genomic DNA"
/isolate="Fec2"
/isolation_source="faeces"
/species="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
/complement(212..862)
/gene="csbA"
/complement(212..862)
/gene="csbA"
/function="necessary for transcription of the csbA
operon"
/codon_start=1

```

```

/transl_table=1
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MPNPFPSIGHITLLITKPSLATALLQHLKOSLALSGKLNIIQ
RLSDISSCCIVLVADDAKKLIKHWODLSKNNLITLLNPDDYPPALILEMPE
HNGFVYADDEBVNAGGLIGRGEYSOKLASYLHSGSNVNSPBDLTHEE
KEILNKIRIGASNIETARSFLISENTYKHLINLPKLAIVKRTQAVSANDNIR"
1623..2078
/gene="csgB"
/locus_tag="csgB"
/function="nucleator"
/codon_start=1
/transl_table=1
/product="nucleation component of curliin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKKKLIFMLLTVLGAPGIAASRYDLASEVFNPLSSGSPN
OAAITGVCTNSAKNRQESKLSLVSGDSQSNRAKDQSGATYFATIAQSGSND
SISQNTGNTAMIIQKSGKNKATITGYGKTAVVAVVROSOMALRVIR"
2119..2571
/gene="csgA"
/locus_tag="csgA"
/locus_tag="csgA"
/locus_tag="csgA"
/locus_tag="csgA"
/function="curli subunit"
/codon_start=1
/transl_table=1
/product="curliin-csgA protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLLQVAAPAIIVSGSLAGSVPMWGSGGGGGGGSSGPESTL
SIVSGVNNALAAQSGARSDITTHQNFPGAGVADGQSDNSTIDLTFNFGKNNATIT
SIVSGKSDITVSGGNNALVNGTASQSVLYVVGQSGNNAITNQ"

```

ORIGIN

Query Match	60.4%;	Score 275.2;	DB 1;	length 2889;
Best Local Similarity	76.8%;	Pred. No. 4.9e-67;		
Matches 350;	Conservative	0;	Mismatches 103;	Indels 3; Gaps 1;

Qy	1	ATGAAACCTTTAAAAGTGGCAGCATTCGACAGAAACGTAAGTTTCTGGGAGAGCTCTGGCT	60
Db	2119	ATGAAACTTTTAAACAAGTGGCAGCATTTTCAGCAAAATCGAGTTTCTGGCAGAGCTCTGGCT	2178
Qy	61	GGCGTCGTTCCACAATGGGGCGGGCGCGCGTATCATTAACGGCGCGCGCAATAGTCCGAC	120
Db	2179	GGTTCTGTTCCGCAATGGGGCGGGCGGG---TGGCGGGCGGGCGGGAGCAGCTCCGGC	2235
Qy	121	CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGGCCTTGCTCGAA	180
Db	2236	CCGGAATGACCCCTGAGCATTTATCAGCAGAGATCAATTAACGCGCCGCTTGCGCTGCAA	2295
Qy	181	AGCGATGCCCGTAAATCTGAAACGACCAATTACCCGAGCGGTTATGTATACGGCGCGCAT	240
Db	2296	AGCGACGTCGTAAATCTGATACGACCACTTATCATCGAATGGCTTTGGTAAACGGCGCAGAC	2355
Qy	241	GTAGGCCAGGGTGGCGGATATAGTACTATTTGAATCTGACTCGAATGTTTCAGAAATAT	300
Db	2356	GTGGGCCAGGGCTCAGATTAACAGCACCATTCATCTGAATCAAAAACGGCTTCAAAAACAAC	2415
Qy	301	GCCACCAATCGACCAAGTGGAACGCTAAAAATCCGATATTAATCTGTCGGGCATTAACGGCGGT	360
Db	2416	GCCACCAATCGATCAAGTGGAACGGCAAAAATTCGACATTACTGTGAGCAGATATGTGTGA	2475
Qy	361	AATAACGCCGCGCTGTGTAATTAATATGATCAGCTGTATTACCCGTGTGTTATACCATGAATG	420
Db	2476	CATAACGCCGCGCATGTGTGAACCAAGACTGCTCCGATTCCAAGGTTCTGTGCAATCAGGTT	2535
Qy	421	GCACATGCAACCAACGCCGCTAACCAAGATTTAA	456
Db	2536	GGTTTGGCAACCAACGCCGCTAACCAAGATTTAA	2571

RESULT	10
CERS15701	
LOCUS	CPR515701
DEFINITION	Citrobacter freundii csb gene, csd gene and csgA gene.
ACCESSION	AJ515701
VERSION	AJ515701.1 GI:31790495
KEYWORDS	csbA gene; csbB gene; csd gene; cullin-csgA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE	Citrobacter freundii
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales; Enterobacteriaceae; Citrobacter.
REFERENCE	1 Zogal,X., Bokranz,W., Nimtz,M. and Romling,U. Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract Infect. Immun. 72 (7), 4151-4158 (2003)
JOURNAL	2 (bases 1 to 2920)
REFERENCE	Romling,U.
AUTHORS	Direct Submission
TITLE	Submitted (12-NOV-2002) Romling U., Microbiology and Tumorigenology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
JOURNAL	
FEATURES	
SOURCE	1..2920

ORIGIN

Query Match 51.0%; Score 232.4; DB 1; Length 2920;
 Best Local Similarity 71.1%; Pred. No. 7.3e-55;
 Matches 324; Conservative 0; Mismatches 126; Indels 6; Gaps 1;

1 ATGAACTTTTAAAGTGGACGATTCGACCAATTCGATTTCTGCGAGTGTCTGCT
 2123 ATGAACTTTTAAAGTGGACGATTCGACCAATTCGATTTCTGCGAGTGTCTGCT 2182
 61 GCGCTGTTTCCACATATGGGGGCGGGGATATCATATACCGCGCGCAATAGTCCGCG
 2183 GGTGTTGTTCCGCAATGGGGCGGGC-----ATCATCATGTGTGTGGCATATATATGGC 2236
 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTGCGTTACGCTGCGCTGCTCTGCA 180
 2237 CCAACTCTTTCACGAGTATCTACCAATAGGTTCAAACTCTGCAAAATGCTCTGCAA 2296
 181 AGCGATCCCGTAAATGTGAAGACCATTAACCGAGCGGTTATGTATAGCGGCGGAT 240
 2297 AGTGATGCGGTAAATCATATGTACCAATCACACAAACGCGCGTGGCAAGGAGCTGTT 2356
 241 GTAGCGCAGGTCGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 2357 GTTGGCCAGGTCGCTGATGACAGTACATTAGCCGTAACAGCGGCTTCCAGAACAGT 2416
 301 GCCACCATGACACGAGTGAAGCTAAATCCGATTTACTGTGCGCAATACGCGCGGT 360
 2417 GCCATATGATCATGATGGAATGCAAAATGCTGATATAGCGTACCCAGTTCGCTGCG 2476
 361 AATAAGCCGCGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 2476
 2477 CGCAACGTCGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
 421 GCACATGCAACCAACGCGCGCTACACAGCCTGACCTCAATGTGCTGATGACAGAGTT 2536
 2537 GCGTTTGTTACCAAGCTACAGCTACCAACTAA 2572

RESULT 11
 SEUS3207 230 bp DNA linear BCT 01-MAY-1996
 LOCUS Salmonella enteritidis SEF17 fimbria (agfa) gene, partial cds.
 DEFINITION US3207
 ACCESSION US3207.1 GI:1293677
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Salmonella enteritidis
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.
 1 (bases 1 to 230)
 Cox, J.M., Eggleston, S. and Woolcock, J.B.
 Virulence of Salmonella enteritidis in chickens correlates with
 colony morphology and expression of SEF17 fimbriae
 Unpublished
 2 (bases 1 to 230)
 Cox, J.M., Eggleston, S. and Woolcock, J.B.
 Direct Submission
 Submitted (01-APR-1996) J.M. Cox, Food Science and Technology, The
 University of New South Wales, Sydney, NSW 2052, Australia
 Location/Qualifiers
 1..230
 /organism="Salmonella enteritidis"
 /mol_type="genomic DNA"
 /strain="SE30"
 /db_xref="taxon:592"
 /note="amplification product"
 1..230
 /gene="agfa"
 <1..>230
 /gene="agfa"
 /codon_start=1
 /transl_table=11
 /product="SEF17 fimbria"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source

ORIGIN

Query Match 50.2%; Score 229; DB 1; Length 230;
 Best Local Similarity 99.6%; Pred. No. 6.7e-54;
 Matches 229; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

88 GGTATTCATTAACGGCGCGCAATAGTCCGCGCGGACCTCAACGTTGATTTATGAG 147
 1 GGTATTCATTAACGGCGCGCAATAGTCCGCGCGGACCTCAACGTTGATTTATGAG 60
 148 TACGTTCCGCTAACGCTGCGCTTGTCTGCAAGGATGCGGTTAATCTGAACGAGC 207
 61 TACGTTCCGCTAACGCTGCGCTTGTCTGCAAGGATGCGGTTAATCTGAACGAGC 120
 208 ATTACCCAGGCGGTTATGTAACGCGCGCGGATGTAAGCGGTCGCGGATTAATGTA 267
 121 ATTACCCAGGCGGTTATGTAACGCGCGCGGATGTAAGCGGTCGCGGATTAATGTA 180
 268 ATTACCCAGGCGGTTATGTAACGCGCGCGGATGTAAGCGGTCGCGGATTAATGTA 317
 181 ATTACCCAGGCGGTTATGTAACGCGCGCGGATGTAAGCGGTCGCGGATTAATGTA 230

RESULT 12
 ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
 LOCUS E.coli cs9G, cs9F, cs9E, cs9D, cs9B, cs9A, and orfC genes.
 DEFINITION X90754.1 GI:1147558
 ACCESSION X90754.1 GI:1147558
 VERSION
 KEYWORDS
 cs9G gene; cs9B gene; cs9D gene; cs9E gene; cs9F gene; cs9A gene;
 orfC gene.
 SOURCE
 ORGANISM
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 1 (bases 1 to 4680)
 Hammar, M., Arngvist, A., Bian, Z., Olsen, A. and Normark, S.
 Expression of two cs9 operons is required for production of
 fibronectin- and congo red-binding curli polymers in Escherichia
 coli K-12
 MOL. Microbiol. 18 (4), 661-670 (1995)
 96414468
 8817489
 2 (bases 1 to 4680)
 Hammar, M.
 Direct Submission
 Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
 Microbiology and Tumourbiology Center, Box 280, S-171 77 Stockholm,
 SWEDEN
 Location/Qualifiers
 1..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:562"
 /map="23.15 minutes"
 1..3140
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="MC4100"
 /db_xref="taxon:562"
 3141..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="W3110"
 /db_xref="taxon:562"
 <1..>4680
 /insertion_seq="IS2 (partial)"
 complement(133..966)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 FEATURES
 source

source
 source
 repeat_region
 gene


```
/gene="csbg"
complement(133. .966)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62277.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translaction="MKRKLFLVAVMLISGCLTAPPEKARPTLMPRAOSYKDLTHLPA
PGKIFVSVYNIODETGFQKPYPAVSFSTAVPOSATMLVTALDSRPFIPLEBQGLA
NLNBERKIPRAOENGCTAIINRPILOSTITANIMSGSITIGESNYSKGVAYRGS
IGADTQYQDLOIAVNLRVNVSTGEILSSVTSKTIISYEQAGVRFIDQRLBEGE
VGYTSNEPVMCLMSALETGVIFLINDIDGMLDWLONKARONDILVKYHMSVPP
S"
complement(993. .1409)
/gene="csbg"
complement(993. .1409)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62278.1"
/db_xref="GI:1147560"
/db_xref="SWISS-PROT:P52104"
/translaction="MKRVHAAVLLMLISPLSWAGTMTFQFRRNPNFGKNPNNAFLNLS
AOQNSYKDPBYNDPFIETPSALDFTQALQSOILGGLSNINTGKRGKRWVTDYIV
DIARDGLOLQNLDRKTGOTSTIQVSGLOKNNSTDF"
complement(1434. .1823)
/gene="csbg"
complement(1434. .1823)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translaction="MKRIKRWIVAAEFLLFAAGNLAAVEVPGLLTDHTVSSIGHDFY
RASFDMKESDVTGNLTINERPSARMSGITTTVNQDVIFFQFLFPLKRDPEKTVVPL
IOTBEALNRROINALLSTGBLADPF"
complement(1828. .2478)
/gene="csbg"
complement(1828. .2478)
/gene="csbg"
complement(1828. .2478)
/note="necessary for transcription of the csbA operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
/protein_id="CAA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translaction="WFEVBSHGHITLLITIKSSLQATALLQHLKQSLATTKLHNTIO
RSJLDISSGIIILDMVEADKLIHYQDTLSRNNNIKILLNTPBDYPRDIENNP
HINGVFYSMEQERVNGLQVLRGECFTQKLASYLITSGNTRVSTESALLTHEE
KEILNKRIGASNNEIARSLPISENVYKTHLYNLFKIAVAGNRTOAVSWANDNLR"
complement(12626)
/note="transcriptional start site for csbD operon"
3140
/note="transcriptional start site for csbA operon"
3233. .3688
/gene="csbg"
3233. .3688
/gene="csbg"
/codon_start=1
/transl_table=1
/product="nucleation component of curli monomers"
/protein_id="CAA62281.1"
```

```
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translaction="MKRKLFLVAVMLISGCLTAPPEKARPTLMPRAOSYKDLTHLPA
PGKIFVSVYNIODETGFQKPYPAVSFSTAVPOSATMLVTALDSRPFIPLEBQGLA
NLNBERKIPRAOENGCTAIINRPILOSTITANIMSGSITIGESNYSKGVAYRGS
IGADTQYQDLOIAVNLRVNVSTGEILSSVTSKTIISYEQAGVRFIDQRLBEGE
VGYTSNEPVMCLMSALETGVIFLINDIDGMLDWLONKARONDILVKYHMSVPP
S"
complement(993. .1409)
/gene="csbg"
complement(993. .1409)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62277.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translaction="MKRKLFLVAVMLISGCLTAPPEKARPTLMPRAOSYKDLTHLPA
PGKIFVSVYNIODETGFQKPYPAVSFSTAVPOSATMLVTALDSRPFIPLEBQGLA
NLNBERKIPRAOENGCTAIINRPILOSTITANIMSGSITIGESNYSKGVAYRGS
IGADTQYQDLOIAVNLRVNVSTGEILSSVTSKTIISYEQAGVRFIDQRLBEGE
VGYTSNEPVMCLMSALETGVIFLINDIDGMLDWLONKARONDILVKYHMSVPP
S"
complement(993. .1409)
/gene="csbg"
complement(993. .1409)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62278.1"
/db_xref="GI:1147560"
/db_xref="SWISS-PROT:P52104"
/translaction="MKRVHAAVLLMLISPLSWAGTMTFQFRRNPNFGKNPNNAFLNLS
AOQNSYKDPBYNDPFIETPSALDFTQALQSOILGGLSNINTGKRGKRWVTDYIV
DIARDGLOLQNLDRKTGOTSTIQVSGLOKNNSTDF"
complement(1434. .1823)
/gene="csbg"
complement(1434. .1823)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translaction="MKRIKRWIVAAEFLLFAAGNLAAVEVPGLLTDHTVSSIGHDFY
RASFDMKESDVTGNLTINERPSARMSGITTTVNQDVIFFQFLFPLKRDPEKTVVPL
IOTBEALNRROINALLSTGBLADPF"
complement(1828. .2478)
/gene="csbg"
complement(1828. .2478)
/gene="csbg"
complement(1828. .2478)
/note="necessary for transcription of the csbA operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
/protein_id="CAA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translaction="WFEVBSHGHITLLITIKSSLQATALLQHLKQSLATTKLHNTIO
RSJLDISSGIIILDMVEADKLIHYQDTLSRNNNIKILLNTPBDYPRDIENNP
HINGVFYSMEQERVNGLQVLRGECFTQKLASYLITSGNTRVSTESALLTHEE
KEILNKRIGASNNEIARSLPISENVYKTHLYNLFKIAVAGNRTOAVSWANDNLR"
complement(12626)
/note="transcriptional start site for csbD operon"
3140
/note="transcriptional start site for csbA operon"
3233. .3688
/gene="csbg"
3233. .3688
/gene="csbg"
/codon_start=1
/transl_table=1
/product="nucleation component of curli monomers"
/protein_id="CAA62281.1"
```

ORIGIN

Query Match 47.7%; Score 217.6; DB 1; Length 4680;
Best Local Similarity 67.3%; Pred. No. 1.2e-50;
Matches 307; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

```
QY 1 ATGAACCTTTAAAGTGGCGCATTCGACGATCGAGTTTCTGCGAGTCTCTGGCT 60
Db 3729 ATGAACCTTTAAAGTGGCGCATTCGACGATCGAGTTTCTGCGAGTCTCTGGCA 3788
QY 61 GGCCTCGTCCACATGAGGGCGGGCGGTATCATACGCGCGCAATAGTTCCGGC 120
Db 3789 GGTGTGTTCTCTAGTACGGGGCGGTATCACCGGTGTGCGGTAAATATAGCGGC 3848
QY 121 CCGACCTGACGTTGAGCATTTATCAGTACGGTTCCGCTAACGTCGCTTCTGCA 180
Db 3849 CCAATCTGACGTCGAACTTACGACGTACGGTGGCGGTAACTTCGACTTCTGCA 3908
QY 181 AGGATCCCGTAAATCGAAGACCATTAACCGAGCGGTTATGGTAACGGCGCGAT 240
Db 3909 ACTATGCCCGTACTTGTGACTTACATTAACCGACATGCGCGGTTATGGTGCAGAT 3968
QY 241 GTAGCCAGGGTGGGATTAATGACTATTGAACCTGACTGAAATGCTTTCAGAAATAT 300
Db 3969 GTTGGTCAGGGCTCAGATGACAGTCAATCGATCGAACCAATGCTTCTGTAACGCG 4028
QY 301 GCCACATCGACGATGGAACGCTAAACCTCCGATATTACTGCGGCAATACGGCGGT 360
Db 4029 GCTACTCTTATGATGATGGAACGGCAAAATTCGAAATGACGGTTAAACAGTTGGTGT 4088
QY 361 AATACCGCGCGCTGTTAATTAATGATCAGCTGGTTACCGGTGTTTACCATGAATATG 420
Db 4089 GCGAACGCTGCTGAGATTGACAGACTGCATCTAACCTTCCTCCGCAACGTACTCAGTT 4148
QY 421 GCACATGCAAAACAAGCCACGCGCTAACCAAGTATTA 456
Db 4149 GCGTTTGTAAACAAGCCGCGCTCATCAGTACTAA 4184
```

```

RESULT 13
AE000205
LOCUS
DEFINITION Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
ACCESSION AE000205 U00096
VERSION AE000205.1 GI:1787265
SOURCE
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Klinkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
TITLE
JOURNAL 9742617
MEDLINE 9278503
PUBMED 2 (bases 1 to 10346)
REFERENCE
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
608-263-7459
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
3 (bases 1 to 10346)
REFERENCE
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
608-263-7459
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
4 (bases 1 to 10346)
REFERENCE
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GenMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
source
1.10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"

repeat_region
promoter
promoter
gene
CDS

protein_bind
protein_bind
protein_bind
promoter
gene
CDS

/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
complement(<1..48)
/feature="IS3"
<1..5
/feature="factor Sigma70; predicted +1 start at 1094689"
5..32
/feature="factor Sigma70; predicted +1 start at 1094716"
69..392
/feature="b1028"
69..392
/feature="b1028"
/feature="orf; unknown"
/feature="orf; This 107 aa ORF is 27 pct identical (0 gaps)
to 58 residues of an approx. 1888 aa protein FAS2_CANAL
SW: P43098"
/codon_start=1
/translation="MEDGQAMRDYLAQDIASPERHOLVNNYAKSEVAPSIIKENIS
LHLRTVVKWTSGLDTFLANKPVPVAFLEISNKPVIGNNLNDPAGVDQQRVHCRKARSD
AVPRK"
275..299
/feature="b1028"
/feature="factor Sigma32; predicted +1 start at 1094983"
389..1375
/feature="ycdY"
/feature="synonym: b1029"
389..1375
/feature="ycdY"
/feature="orf; unknown"
/feature="orf; This 328 aa ORF is 19 pct identical (4 gaps)
to 191 residues of an approx. 288 aa protein Y320_MYCB
SW: P47562"
/codon_start=1
/translation="MPTDITRFRQDRNLIYVAILLIGFYKMGAGFTPESDI
ITMSLMSGLVNLNIFDLKAWAVKCVKXNDIPSWFKKXKXIELELTOPLVNGFS
SLIRYVLVSICISTVNLITLISPLNSDSFTGEGOMLTPKSIITALLILGCVLAINIF
ATLPSKRAVFLGLRFLQEIIDPFESNLSSTPEAKPIMKRLFILVLEVMITLSVL
ATLVKRWLWFEAVYFLTCVPLCIYFYPCRLMHNDPMACDMYFRMGHNK"
1253..1280
/feature="ycdY"
/feature="central position to predicted promoter: -104"
1253..1280
/feature="ycdY"
/feature="No predicted promoter"
/feature="bound moiety="Deor predicted site"
1337..1365
/feature="ycdY"
/feature="factor Sigma70; predicted +1 start at 1096049"
1494..1745
/feature="b1030"
1494..1745
/feature="b1030"
/feature="orf; unknown"
/feature="orf; This 83 aa ORF is 25 pct identical (7 gaps)
to 61 residues of an approx. 616 aa protein ALBU_RABIT SW:
P49065"
/translation="MGGRRCVQYTKKKPVLVSYELFRYGGEGIDLSRSPCGQPAHC
/codon_start=1
/translation="11"
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGRRCVQYTKKKPVLVSYELFRYGGEGIDLSRSPCGQPAHC

```

gene
VHG1SNM1SPVUDPRSGLLIPVCAIHEKKARTFVRL11
1513..1926
/gene="ycdW"
/note="synonym: b1031"
1513..1926
/gene="ycdV"
/function="putative structure; Not classified"
/note="Q137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein R317_YEAST SW:
P05759"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKKSPPYFRSSSSINMAVGGLTTPAPAGSPLTAFVCPPT
SGRLSTGRGFSPPRCVQYTKKKEVLSYEYFPKYGEGGIDSLRSPGQPAHCVHGL
SNM1SPVPRSGLLIPVCAIHEKKARTFVRL11"
complement(2111..2198)
/gene="serX"
/note="synonym: b1032"
complement(2111..2198)
/gene="serX"
/product="tRNA-Ser"
/note="anticodon: GGA; CG Site No. 17878"
/function="tRNA; tRNA"
2269..2296
/note="Factor Sigma70; predicted +1 start at 1096980"
2338..2365
/note="Factor Sigma70; predicted +1 start at 1097049"
2355..2383
/note="Factor Sigma70; predicted +1 start at 1097067"
2393..3370
/gene="ycdW"
/note="synonym: b1033"
2393..3370
/gene="ycdW"
/function="putative enzyme; Not classified"
/note="Q325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPPA_CORGL
SW: P45637"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="MRINFAQRFSGVSMIDIIFYHPTDQWMEALRKAIPOARVRAM
KSGNDSDADYALVMHPPVEMLAGRDILKAVFAGVDSILSKLAHPEMLNPSVLP
LEDTGMEOMOEVAVSQVLAHFRFEDPRYRIONSSHNOPLREYREDPTGILAGYL
GSKVAQSLQTRFPLRCMSRTRKSPGVQSPAGEELSAFLSQCRVILNIPMPBY
GIHQQLLEKLPDCAIYLNLARGVHVEDDLALDGSKAKAMLDVFNREPLPESP
LMQHPRVITIPHVAITRPAVAEYISRTIAQLERGERVCGVDNARGY"
3278..3308
/gene="ycdW"
/note="Factor Sigma70; predicted +1 start at 1097992"
3383..3410
/note="Factor Sigma70; predicted +1 start at 1098094"
3396..3423
promoter

Query Match 47.7%; Score 217.6; DB 1; Length 10346;
Best Local Similarity 67.3%; Pred. NO. 1.2e-50;
Matches 307; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

gene
1 ATGAACCTTTAAAGTGGACGACATTCGACGACATTCGACGACGCTCTGGCT 60
Db 8993 ATGAACCTTTAAAGTGGACGACATTCGACGACATTCGACGACGCTCTGGCA 9052

gene
61 GGGCTGCTTCAACATGGGGGGGGGGGGTATCATTAACGGGGGGAATGTTCCGGC 120
Db 9053 GGTGTGTTCTCTGACGAGGGGGGGGGGTAACACGCTGGGGGTAATTAATGAGGCG 9112

gene
121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTCCGCTAACGCTGGCGCTTCTCGCAA 180

Db 9113 CCAATTCTGATGTAACATTTACAGATGAGCGGTGAGTACCTGACCTGCTGCA 9172

Qy 181 AGCGATGCCGTAATATCGAAACGACATTCACGACGCGTTATGTAAGCGCCGAT 240

Db 9173 ACTGATGCCGTAATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 9232

Qy 241 GTAGGCCAGGGTGGGTAATAGTAACTTGAACGTCGAAATGTTTCAGAATAT 300

Db 9233 GTGTGTCGGGCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9292

Qy 301 GCCACATCGACGACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTAACGTA 360

Db 9293 GCTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9352

Qy 361 AATAACCGCGCGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 420

Db 9353 GCGAACGCTGCTGACGATGACACACATGATGATGATGATGATGATGATGATGATGAT 9412

Qy 421 GCACATGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 456

Db 9413 GCGTTGGTAAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 9448

RESULT 14

LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002

DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).

ACCESSION D90741 AB001340

VERSION D90741.1 GI:1651509

KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB; csgA; ycdB; cys nov; mdog.

SOURCE Escherichia coli K12

ORGANISM Escherichia coli K12

REFERENCE 1 Oshima, T., Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K., Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Saito, N., Sampel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., Yano, M. and Horuchi, T. A 718-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 12.7-28.0 min region on the linkage map DNA Res. 3 (3), 137-155 (1996) 8905232

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

COMMENT

Direct Submission
Submitted (29-JUL-1996) Hirokazu Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan (E-mail:hmori@glc.aisc-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
Collaboration Information:
Project:
The Japan E. coli genome DNA sequencing project
Group:
The Japan E. coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Alba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,

Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Salto, N., Sampel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.
 Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kihori@nibb.ac.jp
 Information operator:
 Name: Hirokazu Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmori@nsl.nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://baw3.aist-nara.ac.jp.
 Location/Qualifiers
 1. 15047
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:83333"
 /map="23.7-24.0 min"
 /clone="Kohara clone #211"
 /note="Nucleotide position 1097101-1112147 from the initiation site of Thra (0 min). This clone is from Kohara lambda minist library."
 <1. 276
 /note="unnamed protein product; ORF_ID:0230#5"
 /codon_start=1
 /transl_table=11
 /protein_id="BA35822.1"
 /db_xref="GI:4062605"
 /translation="DLFSSENALSTPFAKPLMLFTLLIVYEWMTTLVSVALIVE
 WRWFEAYFLCLVPCLIYFFCYCRFLMNDPMACMDYFRMGHFNK"
 1152. 2090
 /note="ORF_ID:0231#1
 similar to SwissProt Accession Number P45637"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical 33.0 kd protein in prob-proA
 intergenic region."
 /protein_id="BA35823.1"
 /db_xref="GI:4062606"
 /translation="WDIIFHPTPTDQWMEALKAIPQARVRAKSGNDSDYALV
 MHPVEMLAGRLKAVFALGAGVSIISKQAHPEMLNPSVPLFLEDTGMGOMQY
 AVSOVLHWRPDDYRIQNSSHWQPLPEYHREDFTIGLAGVLGSKVAOSLQMRP
 PLRCWSPTRKSPWQSPAGRELSAFLSQCVLNLPTPTETVGINOOLLEKLP
 GAYLNLARGVHVEDDLAALDGGKVGKMLDVPNRREPLPESPIMOHPRVTITPHV
 AAITRPAEAVYISRTIAOLEKGERVCGQVDRAGY"
 2145. 2882
 /note="unnamed protein product; ORF_ID:0231#2"
 /codon_start=1
 /transl_table=11
 /protein_id="BA35824.1"
 /db_xref="GI:4062607"
 /translation="MYFVLDLMTVASTHAYSTLSDYIAOKOKGIXLPAITDGPDM
 EDAPHWHFTNMRIPWVVDGVLIRGIEANIKNDGETDCGRFDSIDLITGFHE
 FVEAPHKATNTAMATIASGNVHIIISHRGPKYEIDYKVAEAAAHQVALINNS
 SFHSRKSDEDCREVAARVADGAGVVALGSDSHAFATWGEPECKLIDAVDPPER
 ILNVSPPRLNLPFESRGVAPIAEFADL"
 2906. 3460
 /note="ORF_ID:0231#3
 similar to PIR Accession Number F64035"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical protein H11543"
 /protein_id="BA35825.1"
 /db_xref="GI:4062608"
 /translation="MNEPSTILCRVIGSLYTRQDDPLVPLFTLIRGKLIANWPLQ
 DELLRLRQKSCMTQVSADYNALFTIGDECAVPPYRSAMVEGATEAVERARLISRGNPL

CDS

gene

CDS

gene

CDS

gene

CDS

CDS

ADTPADHICTLLLAASWIEDSTDESEALETLFSEYLLPWCAFLGKVEAHATTPFW
 RTMALITDIALSAMWDELBDSEB"
 3514. 4053
 /note="unnamed protein product; ORF_ID:0231#4"
 /codon_start=1
 /transl_table=11
 /protein_id="BA35826.1"
 /db_xref="GI:4062609"
 /translation="MCSISFPAAFSAIMGNMNLIAITTGILSGIMGVASGIL
 SMAGFLCTAYFACPOGGLKGLAISATILSGVWAMVITIGSALAPHEITIGYITG
 IVAFMLCQIAKOLLSPVPGTFIGACATPAGGQDWKLVPLSLALGIIIGYAMKSGLM
 LAARSATYAREDEIKKA"
 complement (4117. 4950)
 /gene="cs9G"
 complement (4117. 4950)
 /note="cs9G"
 /gene="cs9G"
 complement (4117. 4950)
 /note="ORF_ID:0231#5
 similar to SwissProt Accession Number P52103"
 /codon_start=1
 /transl_table=11
 /product="Curli production assembly/transport component
 cs9G precursor."
 /protein_id="BA35827.1"
 /db_xref="GI:4062610"
 /translation="WQRLFLVAVMLSGCLTAPPKEARPTLMPRAQSYKQLTHLPA
 PRGKTFVSYTINQDETQFKPYPAISNFSIVPOGATMLVTALKDSRWFTPLERQLO
 NULNRKRTIRAAQENGTVAINNRIPLOSLTAANIMVGSIIIGESVSGSGVGARYFG
 ICADYQYLDQIAVNLRVNVSSTGSIISVNTSKTILSYVQGVRFIDYQKLLGE
 VGYTNEPVMCLMSAIBETGVIFLINDGIRGLMDLONKAERONDILVRYRHSVPE
 S"
 complement (4977. 5393)
 /gene="cs9F"
 complement (4977. 5393)
 /note="cs9F"
 /note="ORF_ID:0231#6
 similar to SwissProt Accession Number P52104"
 /codon_start=1
 /transl_table=11
 /product="Curli production assembly/transport component
 cs9F precursor."
 /protein_id="BA35828.1"
 /db_xref="GI:4062611"
 /translation="MRKVAIVLMLISPLSWAGTMTQFPRNPFGNPNNGCALILNS
 AQQNSYKQPSYNDPGETPSALDNFTQAIQSGLININTGKRGKRVYNDYIV
 DIANRGGQQLVNDTRKGTSTIQVSGIQNNSTDF"
 complement (5418. 5807)
 /gene="cs9E"
 complement (5418. 5807)
 /note="cs9E"
 /note="ORF_ID:0231#7
 similar to SwissProt Accession Number P52105"
 /codon_start=1
 /transl_table=11
 /product="Curli production assembly/transport component
 cs9E precursor."
 /protein_id="BA35829.1"
 /db_xref="GI:4062612"
 /translation="MKRTLRWIVAAEFLFAAGNLAAVEVEPGLTDTHTVSSIGDFY
 RASFSKMSDITGLTTRNERSPARWGSWITITVQDVIFQTLFPIKRFEXTVFPAL
 IQTEBALRRQINQALSTGDLAHEF"
 complement (5812. 6462)
 /gene="cs9D"
 complement (5812. 6462)
 /note="cs9D"
 /gene="cs9D"
 complement (5812. 6462)
 /note="ORF_ID:0231#8
 similar to SwissProt Accession Number P52106"
 /codon_start=1
 /transl_table=11
 /product="Probable cs9AB operon transcriptional regulatory
 protein."
 /protein_id="BA35830.1"
 /db_xref="GI:4062613"
 /translation="MNEVHSIHGHTLLITKSSIQATVALLQHLKOSIAITGKLNQIO

gene
 RSLDLSGSIILLDMEDAKKLIHYMODLSRKNNIKILLINTPEDYPRDIENWP
 HINGFVMSDEOERYVNGLOGLRGECEFTOKIASYLLTHSGNRYNSTEALLTHRE
 KEILNKLRIGASNNNEIARSLPISENTVTHLYNLFKIAIVAKRTQAVSMAADNLR"

7217..7672
 /gene="esgb"
 7217..7672

Query Match 47.7%; Score 217.6; DB 1; Length 15047;
 Best Local Similarity 67.3%; Pred. No. 1.2e-50;
 Matches 307; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGCAGATTCGAGCAATCGATGTTCTGCGAGTCTTGCT 60
 Db 7713 ATGAACTTTTAAAGTGCAGATTCGAGCAATCGATGTTCTGCGAGTCTTGCA 7772
 Qy 61 GCGCTGTTCCACATGCGGCGCGCGGTAACTAAACGCGCGGCAATAGTCCGAC 120
 Db 7773 GGTGTTGTTCTCAGTACGCGCGCGCGGTAAACGCGGTGCGGTAAATAGCGGC 7832
 Qy 121 CCGGACTCAAGTTCAGATTTATAGTACGCTTCGCTTCCGCTTCTGCA 180
 Db 7833 CCAATTTCTGAGCTGAACATTTACAGTACGCTGCGGTAACTCTGCACTTGCTTGC 7892
 Qy 181 AGCGATGCCGTTAATCTGAAACGACCATTAACCGAGCGGTTATGTAAACGCGCAT 240
 Db 7893 ACTGATGCCGTTAATCTGAGCTGATTTACCGAGCATGCGGCGGTAAATGTGCAAT 7952
 Qy 241 GTAGCCAGGCTGCGATTAATGATCTAATGAACTGACTCAGATGTTTCAAAATAT 300
 Db 7953 GTTGGTCAGGCTCAGATGACGCTCAATGATCTGACCAACGTTGCGGTAAACAGC 8012
 Qy 301 GCCACATCGACGATGGAACGCTTAAATCCGATATTACTGTGCGCATACGCGCGT 360
 Db 8013 GCTACTCTTGATTCAGTGAACGCGCAAAATCTGAAATGACGTTAAACAGTTCGTTGT 8072
 Qy 361 AATAACGCGCGCTGTTAATTAATGATCAGCTGTTACCCGTTGTTTACCATGAATG 420
 Db 8073 GCGAACGCTGCTGCAAGTTGACCAAGCTCAGTCACTCTCCGTCACGTCAGCTT 8132
 Qy 421 GCACATGCAACACGCGCGCTTACCAAGTATTAA 456
 Db 8133 GCGTTTGGTAAACACGCGCGCTCAGTCACTTAA 8168

RESULT 15
 AX814811 456 bp DNA linear PAT 05-DEC-2003
 LOCUS AX814811
 DEFINITION Sequence 15 from Patent WO03064446.
 ACCESSION AX814811
 VERSION AX814811.1 GI:39104001
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCES
 1 Bjoerck, L., Olsen, A., Wikstroem, M. and Herwald, H.
 Peptides
 Patent: WO 03064446-A 15 07-AUG-2003;
 Hansas Medical Research Aktiebolag (SE)
 Location/Qualifiers
 1..456
 /organism="Escherichia coli"
 /mol_type="unassigned DNA"
 /db_xref="taxon:562"
 1..456
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAE83541.1"
 /db_xref="GI:39104002"
 /translation="MKLLKVAIAIAIVFSSGALAGVVPQYGGGNGHGGGNGNSGPNSE
 LNIYGGGNSALALQTDARNSDLITIOHGGGNGADVGGGSDSSIDLTOGFGNSAT"

CDS

ORIGIN LDONKNSSENTVAKQFGGNGAAYDQTASNSVAVTQVGFNNATAHQ"

Query Match 47.4%; Score 216; DB 6; Length 456;
 Best Local Similarity 67.1%; Pred. No. 3.3e-50;
 Matches 306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGCAGATTCGAGCAATCGATGTTCTGCGAGTCTTGCT 60
 Db 1 ATGAACTTTTAAAGTGAAGCAATTCGACCAATCGATTTCTCGGTAGCGCTTGCA 60
 Qy 61 GCGCTGTTCCACATGCGGCGCGCGGTATATCAATACGCGCGGCAATAGTCCGAC 120
 Db 61 GGTGTTGTTCTCAGTACGCGCGCGCGGTAAACAGCTGCTGCGGTAAATAGCGGC 120
 Qy 121 CCGGACTCAAGTTCAGATTTATAGTACGCTTCGCTTCCGCTTCTGCA 180
 Db 121 CCAATTTCTGAGCTGAACATTTACAGTACGCTGCGGTAACTCTGCACTTGCTTGC 180
 Qy 181 AGCGATGCCGTTAATCTGAAACGACCATTAACCGAGCGGTTATGTAAACGCGCAT 240
 Db 181 ACTGATGCCGTTAATCTGAGCTGATTTACCGAGCATGCGGCGGTAAATGTGCAAT 240
 Qy 241 GTAGCCAGGCTGCGATTAATGATCTAATGAACTGACTCAGATGTTTCAAAATAT 300
 Db 241 GTTGGTCAGGCTCAGATGACGCTCAATGATCTGACCAACGTTGCGGTAAACAGC 300
 Qy 301 GCCACATCGACGATGGAACGCTTAAATCCGATATTACTGTGCGCATACGCGCGT 360
 Db 301 GCTACTCTTGATTCAGTGAACGCGCAAAATCTGAAATGACGTTAAACAGTTCGTTGT 360
 Qy 361 AATAACGCGCGCTGTTAATTAATGATCAGCTGTTACCCGTTGTTTACCATGAATG 420
 Db 361 GCGAACGCTGCTGCAAGTTGACCAAGCTCAGTCACTCTCCGTCACGTCAGCTT 420
 Qy 421 GCACATGCAACACGCGCGCTTACCAAGTATTAA 456
 Db 421 GCGTTTGGTAAACACGCGCGCTCAGTCACTTAA 456

Search completed: March 15, 2004, 22:50:01
 Job time : 1964.17 secs

THIS PAGE BLANK (USPTO)

CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC bacteriophage host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match

Best Local Similarity 100.0%; Score 456; DB 3; Length 456;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGGAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGGAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
QY 61 GGGCTGCTTCCCAATGGGCGGCGGCGATATCAATACGCGCGCAATAGTTCCGCG 120
DB 61 GGGCTGCTTCCCAATGGGCGGCGGCGATATCAATACGCGCGCAATAGTTCCGCG 120
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCGCTGTGCA 180
DB 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCGCTGTGCA 180
QY 181 AGCGATCCCGTAAATCTGAAAGCAATTCACGAGGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATCCCGTAAATCTGAAAGCAATTCACGAGGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCGAGGTCGCGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGTCGCGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACCACTGGAACGCTAAAACTCCGATTTACTGTCGCGCAATACGCGGT 360
DB 301 GCCACCATGACCACTGGAACGCTAAAACTCCGATTTACTGTCGCGCAATACGCGGT 360
QY 361 AATTAACGCGCGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 AATTAACGCGCGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GCACATGCAAAACGCGACGCTAACCAAGTATTA 456
DB 421 GCACATGCAAAACGCGACGCTAACCAAGTATTA 456

```

RESULT 2

AA087467

AA087467 standard; DNA; 456 BP.

AC AA087467;

DT 25-MAR-2003 (revised)

OS 26-JUN-1995 (first entry)

DE Agfa sequence.

XX *Salmonella*; Agfa; vaccine; genetic immunization; ds.

XX *Salmonella*.

XX *Salmonella*.

PH Key Location/Qualifiers
 FT CDS 1..454
 FT /tag= a
 FT /note= "Agfa"

MO9425598-A2.

10-NOV-1994.

26-APR-1994; 94WO-IB000207.

26-APR-1993; 93US-00054452.

(UWV1-) UNIV VICTORIA INNOVATION & DEV CORP.

(KING/) KING J.

Kay WW, Collinson SK, Clouthier SC, Doran JL,

WPI: 1994-358275/44.

P-PSDB; AAR74625.

Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*

strains, vector constructs, or compns. contg. fimbrial type proteins.

Disclosure; Fig 7B; 95pp; English.

The DNA encodes the *Salmonella* Agfa protein. The DNA and isolated

proteins are used in genetic immunization and vaccine compositions,

respectively, to elicit an immune response to *Salmonella* in animals (e.g.

food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN

field.)

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match

Best Local Similarity 88.8%; Score 404.8; DB 2; Length 456;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGGAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGGAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
QY 61 GGGCTGCTTCCCAATGGGCGGCGGCGATATCAATACGCGCGCAATAGTTCCGCG 120
DB 61 GGGCTGCTTCCCAATGGGCGGCGGCGATATCAATACGCGCGCAATAGTTCCGCG 120
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCGCTGTGCA 180
DB 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCGCTGTGCA 180
QY 181 AGCGATCCCGTAAATCTGAAAGCAATTCACGAGGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATCCCGTAAATCTGAAAGCAATTCACGAGGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCGAGGTCGCGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGTCGCGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACCACTGGAACGCTAAAACTCCGATTTACTGTCGCGCAATACGCGGT 360
DB 301 GCCACCATGACCACTGGAACGCTAAAACTCCGATTTACTGTCGCGCAATACGCGGT 360
QY 361 AATTAACGCGCGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 AATTAACGCGCGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GCACATGCAAAACGCGACGCTAACCAAGTATTA 456
DB 421 GCACATGCAAAACGCGACGCTAACCAAGTATTA 456

```

RESULT 3

AAT74142

ID AAT74142 standard; DNA; 456 BP.
 XX
 AC AAT74142;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfa gene.
 XX
 KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..456
 FT /*tag= a
 FT /label= agfa_gene_fragment
 FT /transl_except= (pos:367..369,aa:Pro)
 XX
 PN US5635617-A.
 XX
 PD 03-JUN-1997.
 XX
 PF 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UTVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;
 XX
 DR MPI; 1997-309886/28.
 DR P-PSDB; AAM23570.
 XX
 PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX
 PS Claim 1; Col 19-112; 85pp; English.
 XX
 CC The present sequence represents an isolated agfa gene derived from
 CC Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
 CC diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
 CC family Enterobacteria. It can also be used to provide proteins and
 CC antibodies which can be used for assays. The nucleic acid sequence can be
 CC used to provide probes or primers which can specifically hybridise to
 CC nucleic acid molecules from greater than 9% of Salmonella strains that
 CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
 CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
 CC correct PF field.)
 CC
 XX
 SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
 Query Match 88.8%; Score 404.8; DB 2; Length 456;
 Best Local Similarity 93.0%; Pred. No. 1.1e-120;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
 Oy 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGGGACGCTGCTGCT 60
 Db 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGGGACGCTGCTGCT 60
 Oy 61 GGGCGCTTCCACATGG 120
 Db 61 GGGCGCTTCCACATGG 120
 Oy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGATTCGCTAAACGCTGCGCTTCTGCA 180
 Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGATTCGCTAAACGCTGCGCTTCTGCA 180
 Oy 181 AGCGATGCGCCGCTTAATCTGAAACGACCTTACCCAGAGCGGTTATGTAACGCGCCG 240
 Db 181 AGCGATGCGCCGCTTAATCTGAAACGACCTTACCCAGAGCGGTTATGTAACGCGCCG 240
 Oy 241 GTAGCGCAGGGTGGGATTAATAGTACTGTAAGTGAAGTGAAGTGAAGTGAAGTGA 300

Db 241 GTAGCGCAGGGTGGGATTAATAGTACTGTAAGTGAAGTGAAGTGAAGTGAAGTGA 300
 Oy 301 GCCACATCGACGATGGAAAGCTTAAACCTCGATATTACTGTCGCGCAATACGGCGT 360
 Db 301 GCCACATCGACGATGGAAAGCTTAAACCTCGATATTACTGTCGCGCAATACGGCGT 360
 Oy 361 AATTAAGCCCGGCTGCTTAATATGATCAGCTGTTACCGCTGTTGTTACCATGAATG 420
 Db 361 AATTAAGCCCGGCTGCTTAATATGATCAGCTGTTACCGCTGTTGTTACCATGAATG 420
 Oy 421 GCACATGCAAAACAAGCCGCTAACGATTTAA 456
 Db 421 GCTTTGGCAACACGCCGCTAACGATTTAA 456
 RESULT 4
 AAC64617
 ID AAC64617 standard; DNA; 456 BP.
 XX
 AC AAC64617;
 XX
 DT 26-FEB-2001 (first entry)
 DT
 XX
 DE Salmonella enteritidis Agfa DNA sequence SEQ ID NO:1.
 DE
 KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UTVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX
 DR MPI; 2000-672631/65.
 DR P-PSDB; AAB36341.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 134; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SfF17/TFP) nucleation dependent
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising: (2)
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (3)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP, 117 A, 112 C, 122 G, 105 T, 0 U, 0 Other;

Query Match Best Local Similarity 88.8%; Score 404.8; DB 3; Length 456;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 60
DB 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 60
QY 61 GGGCTGCTTCCCAATGGGGCGGGCGGCTTAATCAATACGGCGCGCAATAGTCCGGC 120
DB 61 GGGCTGCTTCCCAATGGGGCGGGCGGCTTAATCAATACGGCGCGCAATAGTCCGGC 120
QY 121 CCGGACTCAAGCTTGAAGCATTTATCATGATCGGTTCCGCTAACGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTGAAGCATTTATCATGATCGGTTCCGCTAACGCTGCTCTGCA 180
QY 181 AGCGATGCGCCGTTAATCTGAAACGACCATTCACGAGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATGCGCCGTTAATCTGAAACGACCATTCACGAGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCCAGGGTGGCGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATTCGACGATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGCGT 360
DB 301 GCCACATTCGACGATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGCGT 360
QY 361 AATAACCGCGGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 AATAACCGCGGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GCACATGCAACACGCGCGGCTTACGAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 456
DB 421 GGTTCGCAACACGCGCGGCTTACGAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 456

```

RESULT 5
 ID AAC64623 standard; DNA; 456 BP.
 XX AC AAC64623;
 XX

26-FEB-2001 (first entry)
 DE AGFA::PT3#2 DNA sequence SEQ ID NO:13.
 XX

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX

Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX

MO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX

05-APR-2000; 2000MO-CA000356.
 XX

05-APR-1999; 99US-0127888P.
 XX

(UYVI-) UNIV VICTORIA.
 XX

White AP, Doran JL, Collison SK, Kay WW;
 XX
 PI

XX WPI; 2000-672631/65.
 DR P-PSDB; AAB36347.
 DR

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX

Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAI) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (It is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP, 118 A, 109 C, 121 G, 108 T, 0 U, 0 Other;

Query Match Best Local Similarity 82.5%; Score 376; DB 3; Length 456;
 Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 60
DB 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 60
QY 61 GGGCTGCTTCCCAATGGGGCGGGCGGCTTAATCAATACGGCGCGCAATAGTCCGGC 120
DB 61 GGGCTGCTTCCCAATGGGGCGGGCGGCTTAATCAATACGGCGCGCAATAGTCCGGC 120
QY 121 CCGGACTCAAGCTTGAAGCATTTATCATGATCGGTTCCGCTAACGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTGAAGCATTTATCATGATCGGTTCCGCTAACGCTGCTCTGCA 180
QY 181 AGCGATGCGCCGTTAATCTGAAACGACCATTCACGAGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATGCGCCGTTAATCTGAAACGACCATTCACGAGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCCAGGGTGGCGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATTCGACGATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGCGT 360
DB 301 GCCACATTCGACGATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGCGT 360
QY 361 AATAACCGCGGCTGTTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 CTGGTACCGGTGTGTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GCACATGCAACACGCGCGGCTTACGAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 456
DB 421 GGTTCGCAACACGCGCGGCTTACGAGCATTCGAGCAATGTGTTCTGCGAGTCTCTGCT 456

```

Db 421 GGTITGGCAACAAGCCAGCTAACCAATTATTA 456

RESULT 6
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#5 DNA sequence SEQ ID NO:19.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36350.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAP) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CagA and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbria protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
Best Local Similarity 87.1%; Pred. No. 1.1e-106;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGCAGCATTTCCGACCAATCGTAGTTTCTGGCAGTCTGGCT 60
Db 1 ATGAACCTTTTAAAGTGGCAGCATTTCCGACCAATCGTAGTTTCTGGCAGTCTGGCT 60
QY 61 GGCCTCGTTCCACAATGAGGCGGCGGTAATCATTAACGGCGGCAATAGTTCCGGC 120
Db 61 GGCCTCGTTCCACAATGAGGCGGCGGTAATCATTAACGGCGGCAATAGTTCCGGC 120
QY 121 CCGACTCAACGTTGAGCATTTATCACTACGTTCCGCTAACCGTGGCTTCTCGCA 180
Db 121 CCGACTCAACGTTGAGCATTTATCACTACGTTCCGCTAACCGTGGCTTCTCGCA 180
QY 181 AGCGATCCCGTAAATGTGAAGACCAATTACCGAGCGGTTATGTGAACGGCGGAT 240
Db 181 AGCGATCCCGTAAATGTGAAGACCAATTACCGAGCGGTTATGTGAACGGCGGAT 240
QY 241 GTAGCCAGGAGTGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
Db 241 GTAGCCAGGAGTGGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 300
QY 301 GCCACATGACGACGAGTGAACGCTAAACTCCGATTTAATCTGCGCAATACGGCGT 360
Db 301 GCCACATGACGAGTGAACGCTAAACTCCGATTTAATCTGCGCAATACGGCGT 360
QY 361 AATTAACCGCGCGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
Db 361 AATTAACCGCGCGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 420
QY 421 GCACATGCAACAAACGCGCAGGCTAACCAAGTATTA 456
Db 421 GGTITGGCAACAAGCCAGCTAACCAATTATTA 456

RESULT 7
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#4 DNA sequence SEQ ID NO:17.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX
DR WPI; 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAAP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of a recombinant AgfA
 CC protein containing a replacement segment or segments of a recombinant amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;
 Query Match 77.9%; Score 355.2; DB 3; Length 456;
 Best Local Similarity 86.2%; Pred. No. 1.3e-104;
 Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGAGCATTCGACGAACTGTAAGTTCTGAGCTCTGCT 60
 DB 1 ATGAACTTTTAAAGTGAGCATTCGACGAACTGTAAGTTCTGAGCTCTGCT 60
 QY 61 GGGCTGCTTCCACATGGGGGGGGGGGGGGGATATCATAGCGGGGCAATAGTTCCGGC 120
 DB 61 GGGCTGCTTCCACATGGGGGGGGGGGGGGGATATCATAGCGGGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGCTTGAGCATTTTATCAGTACGGTCCGCTAACCGCTGCTCTGCA 180
 DB 121 CCGGACTCAAGCTTGAGCATTTTATCAGTACGGTCCGCTAACCGCTGCTCTGCA 180
 QY 181 AGCGATGCCGCTAAATCTGAACGACATTCACCGAGCGGTATGTTAAGGCACTGTCGA 180
 DB 181 AGCGATGCCGCTAAATCTGAACGACATTCACCGAGCGGTATGTTAAGGCACTGTCGA 180
 QY 241 GTAGGCGACGGTGGCGATTAATGTAATCTTAAGCTGACGATGAGATGTTTCAAAATAT 240
 DB 241 GTAGGCGACGGTGGCGATTAATGTAATCTTAAGCTGACGATGAGATGTTTCAAAATAT 240
 QY 301 GGCACCATCGACCAATGGAAGCTGCGATTAATCTGCGGCAATACGCGGT 300
 DB 301 GGCACCATCGACCAATGGAAGCTGCGATTAATCTGCGGCAATACGCGGT 300
 QY 361 AATAAGCGGGCGGTGTTAATTAATGATGACGCTGTTACCGGTGTTTCAATGAATG 360
 DB 361 AATAAGCGGGCGGTGTTAATTAATGATGACGCTGTTACCGGTGTTTCAATGAATG 360
 QY 421 GCACATGCAAAACGACGCGCTAACGATTA 456
 DB 421 GGTGTTGCAAAACGACGCGCTAACGATTA 456
 XX 421 GGTGTTGCAAAACGACGCGCTAACGATTA 456

RESULT 8
 AAC64628
 ID AAC64628 standard; DNA; 456 BP.
 XX AAC64628;
 AC AAC64628;
 XX AAC64628;
 DT 26-FEB-2001 (first entry)
 XX

DE AgfA::PT3#7 DNA sequence SEQ ID NO:23.
 XX *Salmonella*; agfA; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen; ds.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 PD 12-OCT-2000.
 PF 05-APR-2000; 2000WO-CN000356.
 PR 05-APR-1999; 99US-0127888P.
 PA (UWI-) UNIT VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay W;
 DR WPI; 2000-672631/65.
 DR P-PSDB; AAB36352.
 PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAAP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of a recombinant AgfA
 CC protein containing a replacement segment or segments of a recombinant amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;
 Query Match 77.5%; Score 353.6; DB 3; Length 456;
 Best Local Similarity 86.0%; Pred. No. 4.4e-104;
 Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGAGCATTCGACGAACTGTAAGTTCTGAGCTCTGCT 60
 DB 1 ATGAACTTTTAAAGTGAGCATTCGACGAACTGTAAGTTCTGAGCTCTGCT 60
 QY 61 GGGCTGCTTCCACATGGGGGGGGGGGGGGGATATCATAGCGGGGCAATAGTTCCGGC 120
 DB 61 GGGCTGCTTCCACATGGGGGGGGGGGGGGGATATCATAGCGGGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGCTTGAGCATTTTATCAGTACGGTCCGCTAACCGCTGCTCTGCA 180
 DB 121 CCGGACTCAAGCTTGAGCATTTTATCAGTACGGTCCGCTAACCGCTGCTCTGCA 180

Query Match	Similarity	Score	DB 3	Length	456
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.6	DB 3	Length 456	
Best Local	Similarity	86.0%	Pred. No. 4.4e-104		
Matches	392	Conservative	0	Mismatches	64
				Indels	0
				Gaps	0
Sequence	456 BP	114 A	108 C	123 G	111 T
				0 U	0 Other
Query Match	77.5%	Score 353.			

PD 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 XX (UNIV-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR P-PSDB; AAB36354.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;
 Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 1.5e-103;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGGCGCATTCGAGCAATCGTATGTTTGGCAGTCTGGCT 60
 DB 1 ATGAACCTTTAAAGGCGCATTCGAGCAATCGTATGTTTGGCAGTCTGGCT 60
 QY 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATAGTTCGGC 120
 DB 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATAGTTCGGC 120
 QY 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATAGTTCGGC 120
 DB 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATAGTTCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAAACGCTTCGCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAAACGCTTCGCTGCA 180
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAAACGCTTCGCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAAACGCTTCGCTGCA 180
 QY 181 AGGATGCGCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTTAAACGCGCGCAT 240
 DB 181 AGGATGCGCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTTAAACGCGCGCAT 240
 QY 181 AGGATGCGCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTTAAACGCGCGCAT 240
 DB 181 AGGATGCGCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTTAAACGCGCGCAT 240
 QY 241 GTAGGCCAGGCGGTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 300
 DB 241 GTAGGCCAGGCGGTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 300
 QY 241 GTAGGCCAGGCGGTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 300
 DB 241 GTAGGCCAGGCGGTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 300
 QY 301 GCACCATTCGACCAAGTGAGCAAGCTTAAACCGGATTTACTGCGCCCAATACGCGCGT 360
 DB 301 GCACCATTCGACCAAGTGAGCAAGCTTAAACCGGATTTACTGCGCCCAATACGCGCGT 360

DB 301 GCACCATTCGACCAAGTGAGCAAGCTTAAACCGGATTTACTGCGCCCAATACGCGCGT 360
 QY 361 AATAAGCGCGCGGTGTTTATTTATGATCAGTGTTCCTGTTGTTTACCATGAAATG 420
 DB 361 AATAAGCGCGCGGTGTTTATTTATGATCAGTGTTCCTGTTGTTTACCATGAAATG 420
 QY 421 GCACATGCAACACAGCGGCTTACCAAGTATTA 456
 DB 421 GCTTTGCAACAGCGGCTTACCAAGTATTA 456

RESULT 11
 ID AAC64630 standard; DNA; 456 BP.
 XX AAC64630;
 AC
 XX
 XX
 DT 26-FEB-2001 (first entry)
 XX
 XX
 DE Agfa::PT39 DNA sequence SEQ ID NO:27.
 XX
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN WO200060102-A2.
 XX
 XX
 PD 12-OCT-2000.
 XX
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 XX
 PA (UNIV-) UNIV VICTORIA.
 XX
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX WPI; 2000-672631/65.
 DR P-PSDB; AAB36354.
 XX
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP, 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 4.8e-103;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGCTCTGCT 60
QY 61 GGCTTCCTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 61 GGCTTCCTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCGCCAGGCTGGCATATGATGATCTGATGCTGCTGCTGCTGCTGCT 300
DB 241 GTAGCGCCAGGCTGGCATATGATGATCTGATGCTGCTGCTGCTGCTGCT 300
QY 301 GCCGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 360
DB 301 GCCGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 360
QY 361 AATAAGCCCGGCTGTTAATTAATGATCAAGTCTGCTGCTGCTGCTGCT 420
DB 361 AATAAGCCCGGCTGTTAATTAATGATCAAGTCTGCTGCTGCTGCTGCT 420
QY 421 GCACATGCAAAACGCGGCGGCTTAACGATTAATTA 456
DB 421 GCACATGCAAAACGCGGCGGCTTAACGATTAATTA 456
QY 421 GGTTCGCAACGCGGCGGCTTAACGATTAATTA 456
DB 421 GGTTCGCAACGCGGCGGCTTAACGATTAATTA 456
```

RESULT 12

AAC64627 ID AAC64627 standard; DNA; 456 BP.

AC AAC64627;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#6 DNA sequence SEQ ID NO:21.

KM Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;

KX vaccine; immune response; immunogen; ds.

OS Salmonella enteritidis.

OS Escherichia coli.

PN WO20060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000MO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UVT-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR P-PSDB; AAB36351.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/TFP) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, GsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SO Sequence 456 BP, 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 4.8e-103;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGCTCTGCT 60
QY 61 GGCTTCCTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 61 GGCTTCCTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCGCCAGGCTGGCATATGATGATCTGATGCTGCTGCTGCTGCTGCT 300
DB 241 GTAGCGCCAGGCTGGCATATGATGATCTGATGCTGCTGCTGCTGCTGCT 300
QY 301 GCCGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 360
DB 301 GCCGACTCAACGTTGAGCATTTATCAAGTCCGTTACGCTGCTGCTGCA 360
QY 361 AATAAGCCCGGCTGTTAATTAATGATCAAGTCTGCTGCTGCTGCTGCT 420
DB 361 AATAAGCCCGGCTGTTAATTAATGATCAAGTCTGCTGCTGCTGCTGCT 420
QY 421 GCACATGCAAAACGCGGCGGCTTAACGATTAATTA 456
DB 421 GCACATGCAAAACGCGGCGGCTTAACGATTAATTA 456
QY 421 GGTTCGCAACGCGGCGGCTTAACGATTAATTA 456
DB 421 GGTTCGCAACGCGGCGGCTTAACGATTAATTA 456
```


RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 DNA sequence SEQ ID NO:29.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
DR P-FSDB; AAB36355.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising Agfa
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 76.1%; Score 34.72; DB 3; Length 456;
Best Local Similarity 85.1%; Pred No. 5.3e-102;
Matches 388; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
OY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGATCGAGTTCGTGCGAGTCTGGCT 60
XX

Db 1 ATGAACCTTTAAAGTGGCAGCATTCGACGATCGAGTTCGTGCGAGTCTGGCT 60
OY 61 GGGCGTGTTCACATGAGGGGGGGGGGGGATATCATATACGCGCGGCAATATTCGGC 120
Db 61 GGGCGTGTTCACATGAGGGGGGGGGGGGATATCATATACGCGCGGCAATATTCGGC 120
OY 121 CCGACTCAAGCTTACATTTATGAGTGTTCGGCTTACGCTGCGCTTCTGCA 180
Db 121 CCGACTCAAGCTTACATTTATGAGTGTTCGGCTTACGCTGCGCTTCTGCA 180
OY 181 AGCGATCCCGTAAATCTGAAACGACATTAACGAGCGGTTATGTATAGCGCGCGAT 240
Db 181 AGCGATCCCGTAAATCTGAAACGACATTAACGAGCGGTTATGTATAGCGCGCGAT 240
OY 241 GTAGGCGAGGTCGCGATATATGATATTAATGATATTAATGATATTAATGATAT 300
Db 241 GTAGGCGAGGTCGCGATATATGATATTAATGATATTAATGATATTAATGATAT 300
OY 301 GCCACCATGACGACATGCGAACCGCTAATAAATCCGATATTAATGATATGAGGCGGT 360
Db 301 GCCACCATGACGACATGCGAACCGCTAATAAATCCGATATTAATGATATGAGGCGGT 360
OY 361 AATTAACCGCGCGCTGTTAATTAATGATATGATATGATATGATATGATATGATAT 420
Db 361 AATTAACCGCGCGCTGTTAATTAATGATATGATATGATATGATATGATATGATAT 420
OY 421 GCACATGCAACACGCGACCGGCTTAACCGATTTAA 456
Db 421 GGTTTGGCAACACGCGACCGGCTTAACCGATTTAA 456

RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
AC AAQ73066;
XX
DT 27-AUG-2003 (revised)
DT 25-WAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
XX
XX Salmonella; Agfa; vaccine; genetic immunization; ds.
OS Salmonella enteritidis.
XX
FH Key Location/Qualifiers
FH CDS 1..359
FT /tag= a
FT /note= "Agfa"
FT misc_feature 37..60
FT /tag= b
FT /note= "TAF5 primer (pair with TAF6)"
FT misc_feature 52..69
FT /tag= b
FT /note= "TAF3 primer (pair with TAF4)"
FT misc_feature Complement(103..129)
FT /tag= e
FT /note= "TAF6 primer (pair with TAF5)"
FT Complement(292..402)
FT /tag= c
FT /note= "TAF4 primer (pair with TAF3)"
XX
PN WO9425598-A2.
XX
XX 10-NOV-1994.
PD 26-APR-1994; 94WO-IB000207.
PR 26-APR-1993; 93US-00054452.
XX

PA (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI: 1994-358275/44.
DR P-PSDB; AAR62761.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compans. cong. fimbrial type proteins.
XX
PS Disclosure: Fig 7A; 95pp; English.
XX
CC The DNA encodes the Salmonella enteritidis 27655-3b *TrpH* mutant strain
CC *agfA* gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to Salmonella in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 69.1%; Score 315.2; DB 2; Length 361;
Best Local Similarity 99.1%; Pred.No. 1.1e-91;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 64 GTCGTTCCACAAATGGGGGGGGGGGGGCGGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 123
DB 1 GTCGACCAACAGTGGGGGGGGGGGGCGGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 60

QY 124 GACTCAACGTTGAGCATTATTCAGTACGCTCCGCTTAACGGCGGCTTCTCTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTATTCAGTACGCTCCGCTTAACGGCGGCTTCTCTGCAAGC 120

QY 184 GATCCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCGCATGTA 243
DB 121 GATCCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCGCATGTA 180

QY 244 GGCAGGGTGGGATTAATGACTTGAATGACTGATCGAATGTTTCAAGAAATATATGCC 303
DB 181 GGCAGGGTGGGATTAATGACTTGAATGACTGATCGAATGTTTCAAGAAATATATGCC 240

QY 304 ACCATCGACGATGGAACGCTAAACAACTCCGATATTACTGCGGCAATACGGCGGTAT 363
DB 241 ACCATCGACGATGGAACGCTAAACAACTCCGATATTACTGCGGCAATACGGCGGTAT 300

QY 364 AACGCGCGCTGTTAATTA 383
DB 301 AACGCGCGCTGTTAATCA 320

RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
XX AAT74141;
AC
XX 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
XX *Salmonella enteritidis* 27655-3b *TrpH* mutant *agfA* gene fragment.
DE
XX Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody; ds.
XX
XX *Salmonella enteritidis*.
OS
XX
FH Key 1. .360 Location/Qualifiers
FT CDS /tag= a
FT primer_bind /label= *agfA*_gene_fragment
FT 16. .60
FT primer_bind /tag= b
FT /label= Primer_TAF5

FT primer_bind 52. .69
FT /tag= c
FT /label= Primer_TAF3
FT complement(103. .128)
FT primer_bind /tag= d
FT /label= Primer TAF6
FT complement(294. .312)
FT /tag= e
FT /label= Primer_TAF4
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX
XX WPI: 1997-309886/28.
XX
XX P-PSDB; AAM23569.
XX
XX Isolated *Salmonella* gene *agfA* - used for diagnosis of *Salmonella* or
XX enteropathogenic bacteria of the Enterobacteria family.
XX
XX Claim 1; Col 107-110; 85pp; English.
XX
XX The present sequence represents an isolated *agfA* gene fragment derived
XX from *Salmonella enteritidis* 27655-3b *TrpH* mutant strain. The nucleic
XX acid can be used to provide diagnostic assays for *Salmonella* and/or
XX enteropathogenic bacteria of the family Enterobacteria. It can also be
XX used to provide proteins and antibodies which can be used for assays. The
XX nucleic acid sequence can be used to provide probes or primers which can
XX specifically hybridise to nucleic acid molecules from greater than 99% of
XX *Salmonella* strains that are pathogenic to warm-blooded animals relative
XX to nucleic acid molecules from virtually all other microbial organisms.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 69.1%; Score 315.2; DB 2; Length 361;
Best Local Similarity 99.1%; Pred.No. 1.1e-91;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 64 GTCGTTCCACAAATGGGGGGGGGGGGGCGGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 123
DB 1 GTCGACCAACAGTGGGGGGGGGGGGCGGTATCATTAACGGCGGCGCAATAGTTCCGGCCG 60

QY 124 GACTCAACGTTGAGCATTATTCAGTACGCTCCGCTTAACGGCGGCTTCTCTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTATTCAGTACGCTCCGCTTAACGGCGGCTTCTCTGCAAGC 120

QY 184 GATCCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCGCATGTA 243
DB 121 GATCCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCGCATGTA 180

QY 244 GGCAGGGTGGGATTAATGACTTGAATGACTGATCGAATGTTTCAAGAAATATATGCC 303
DB 181 GGCAGGGTGGGATTAATGACTTGAATGACTGATCGAATGTTTCAAGAAATATATGCC 240

QY 304 ACCATCGACGATGGAACGCTAAACAACTCCGATATTACTGCGGCAATACGGCGGTAT 363
DB 241 ACCATCGACGATGGAACGCTAAACAACTCCGATATTACTGCGGCAATACGGCGGTAT 300

QY 364 AACGCGCGCTGTTAATTA 383
DB 301 AACGCGCGCTGTTAATCA 320

Search completed: March 15, 2004, 17:51:56

Thu Mar 18 12:27:55 2004

us-09-543-407-11.png

Page 12

Job time : 246.684 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 / Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456

Sequence: 1 atgaactcttaaaagtggc.....ccacgctaacagattataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCtUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	404.8	88.8	456	1 US-08-233-788A-58	Sequence 58, Appl
2	315.2	69.1	361	1 US-08-233-788A-56	Sequence 56, Appl
3	34	7.5	1344	3 US-09-120-927-1	Sequence 1, Appl
4	34	7.5	1344	4 US-09-431-614-5	Sequence 5, Appl
5	33.4	7.3	7766	4 US-09-125-619-3	Sequence 3, Appl
6	33.4	7.3	14770	4 US-09-220-132-30	Sequence 30, Appl
7	32.2	7.1	1122	4 US-09-489-039A-3211	Sequence 3211, Ap
8	31.2	6.8	1491	6 5486473-3	Patent No. 5486473
9	31.2	6.8	2436	4 US-09-540-236-492	Sequence 492, App
10	31.2	6.8	65792	4 US-09-596-002-31	Sequence 1, Appl
11	30.8	6.8	2547	3 US-08-508-761B-1	Sequence 31, Appl
12	30.6	6.7	4403765	3 US-09-103-840A-1	Sequence 2, Appl
13	30.6	6.7	4411529	3 US-09-103-840A-1	Sequence 1, Appl
14	30	6.6	1341	4 US-09-350-756-2	Sequence 2, Appl
15	29.8	6.5	1008	4 US-09-252-991A-10904	Sequence 10904, A
16	29.8	6.5	1077	4 US-09-252-991A-10818	Sequence 10818, A
17	29.8	6.5	1521	4 US-09-252-991A-10865	Sequence 10865, A
18	29.6	6.5	972	4 US-09-328-352-2055	Sequence 2055, Ap
19	29.4	6.4	423	4 US-09-252-991A-7641	Sequence 7641, Ap
20	29.4	6.4	1252	4 US-09-561-756-29	Sequence 29, Appl
21	29.4	6.4	1252	4 US-09-227-721-29	Sequence 29, Appl
22	29.4	6.4	1252	4 US-09-954-697-39	Sequence 29, Appl
23	29.4	6.4	1939	4 US-08-961-527-310	Sequence 310, App
24	29.4	6.4	3494	3 US-09-135-802-200	Sequence 200, App
25	29.4	6.4	3494	4 US-09-659-786-200	Sequence 200, App
26	29.4	6.4	3494	4 US-09-023-655-1006	Sequence 1006, Ap
27	29.4	6.4	3660	4 US-09-253-991A-12269	Sequence 12269, A

28	29.4	6.4	536165	4 US-09-214-808-1	Sequence 1, Appl
29	29.2	6.4	4403765	3 US-09-103-840A-2	Sequence 2, Appl
30	29.2	6.4	4411529	3 US-09-103-840A-1	Sequence 1, Appl
31	29	6.4	534	4 US-09-489-039A-2195	Sequence 2195, Ap
32	29	6.4	1200	4 US-09-134-000C-1704	Sequence 1704, Ap
33	29	6.4	1818	4 US-09-489-039A-2162	Sequence 2162, Ap
34	29	6.4	13794	4 US-08-956-111E-54	Sequence 54, Appl
35	28.8	6.3	654	4 US-09-252-991A-4506	Sequence 4506, Ap
36	28.8	6.3	720	4 US-09-328-352-354	Sequence 354, App
37	28.8	6.3	1662	1 US-08-565-386-2	Sequence 1, Appl
38	28.8	6.3	4529	1 US-08-565-386-1	Sequence 1, Appl
39	28.8	6.3	8906	4 US-09-027-169-5	Sequence 5, Appl
40	28.6	6.3	432	4 US-09-252-991A-4473	Sequence 4473, Ap
41	28.4	6.2	1026	3 US-07-751-881B-24	Sequence 24, Appl
42	28.4	6.2	4188	3 US-07-751-881B-2	Sequence 2, Appl
43	28.4	6.2	4242	3 US-07-705-490-2	Sequence 2, Appl
44	28.4	6.2	5222	3 US-07-751-891B-23	Sequence 23, Appl
45	28.2	6.2	357	4 US-09-489-039A-1118	Sequence 1118, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 4.5e-111;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTAAAGTGGACATTCGACGAATCGAGTTTCTGGAGTGTCTGGCT
DB 1 ATGAACTTTAAAGTGGACATTCGACGAATCGAGTTTCTGGAGTGTCTGGCT
QY 61 GGCCTGCTTCCCAATGGGGGGGGGGGGGATATCAATACGCGCGCAATAGTTCCGGC
DB 61 GGCCTGCTTCCCAATGGGGGGGGGGGGGATATCAATACGCGCGCAATAGTTCCGGC
QY 121 CCGGATCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCA
DB 121 CCGGATCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCA
QY 181 AGCGATGCGCTTAATCTGAAGACCAATTCGCAAGCGGTTATGTATCGCGCGCAT
DB 181 AGCGATGCGCTTAATCTGAAGACCAATTCGCAAGCGGTTATGTATCGCGCGCAT
QY 241 GTAGCGCGGCTCGATTAATAGTACTATTAATGACTCAGAAATGTTTCAAAATAT
DB 241 GTAGCGCGGCTCGATTAATAGTACTATTAATGACTCAGAAATGTTTCAAAATAT
QY 301 GCCACCATGACCAAGTGAACGCTAAATCCGATTTACTGCGCCAAATCGCGGCT
DB 301 GCCACCATGACCAAGTGAACGCTAAATCCGATTTACTGCGCCAAATCGCGGCT
QY 361 AATAACGCGCGCTGTTATTAATGATGATGATGATGATGATGATGATGATGATGAT
DB 361 AATAACGCGCGCTGTTATTAATGATGATGATGATGATGATGATGATGATGATGAT
QY 421 GCACATGCAACCAACGCGCGCTAACGATTTAA 456
DB 421 GGTATTGCAACCAACGCGCGCTAACGATTTAA 456

```

RESULT 2

```

US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635517
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Cloutier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043,403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
; US-08-233-788A-56

```

```

Query Match
Best Local Similarity 69.1%; Score 315.2; DB 1; Length 361;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 64 GTCTGCAACATGCGCGCGCGCGCGGATCATTAACGCGCGCAATAGTTCCGGCCG
DB 1 GTCTGCAACATGCGCGCGCGCGCGGATCATTAACGCGCGCAATAGTTCCGGCCG
QY 124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCAAGC
DB 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCAAGC
QY 184 GATGCGCGTAAATCTGAACGCACTTAACCAAGCGGTTATGTTACGCGCGCATGTA
DB 121 GATGCGCGTAAATCTGAACGCACTTAACCAAGCGGTTATGTTACGCGCGCATGTA
QY 244 GGCAGGCTGCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT
DB 181 GGCAGGCTGCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT
QY 304 ACCATGCAACAGTGAACGCTAAATCCGATTTACTGCGCCAAATCGCGGCTAT
DB 241 ACCATGCAACAGTGAACGCTAAATCCGATTTACTGCGCCAAATCGCGGCTAT
QY 364 AACGCGCGCTGTTAATTA 383
DB 301 AACGCGCGCTGTTAATTA 320

```

RESULT 3

```

US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICTOR FROM
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

```

TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-120-927-1

Query Match 7.5%; Score 34; DB 3; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGGCTGGCGCTTCCCAATGGGGCGCGCGGTAATCATACGGCGGGCAAT 111
DB 209 GCTCCGGTTGGCGATTACCTGATTTGGCGATGACGATGCTTAACAGTTCCGCCAAT 150
QY 112 AGTTCGGCGCGGAGCTCAAGCTGAGCATTTATGAGTACGGTTCGGCTAAGCGCTG 171
DB 149 AATTGACCATTTCTCATGTTGTCGATGATGGGTTGTCGCCCAACGAGAAATTT 90

QY 172 GC 173
DB 89 GC 88

RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:

APPLICANT: Wei, Zhong-Min
APPLICANT: Schading, Richard L.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
FILE REFERENCE: 21829/41 (EBC-003)
CURRENT APPLICATION NUMBER: US/09/431,614
CURRENT FILING DATE: 1999-11-02
EARLIER APPLICATION NUMBER: 60/107,243
EARLIER FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1344
TYPE: DNA
ORGANISM: *Erwinia amylovora*
US-09-431-614-5

Query Match 7.5%; Score 34; DB 4; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGGCTGGCGCTTCCCAATGGGGCGCGCGGTAATCATACGGCGGGCAAT 111
DB 209 GCTCCGGTTGGCGATTACCTGATTTGGCGATGACGATGCTTAACAGTTCCGCCAAT 150
QY 112 AGTTCGGCGCGGAGCTCAAGCTGAGCATTTATGAGTACGGTTCGGCTAAGCGCTG 171
DB 149 AATTGACCATTTCTCATGTTGTCGATGATGGGTTGTCGCCCAACGAGAAATTT 90
QY 172 GC 173
DB 89 GC 88

RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
APPLICANT: NORRIS, STEVEN J.
APPLICANT: JING-REN, ZHANG

APPLICANT: HARDHAM, JOHN M.
APPLICANT: HOWELL, JERRILYN K.
APPLICANT: BARBOUR, ALAN G.
APPLICANT: WEINSTOCK, GEORGE M.
TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
FILE REFERENCE: USH:234
CURRENT APPLICATION NUMBER: US/09/125,619
CURRENT FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 7766
TYPE: DNA
ORGANISM: *Borrelia burgdorferi*
FEATURE:
NAME/KEY: modified_base
LOCATION: (127)
OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 7.3%; Score 33.4; DB 4; Length 7766;
Best Local Similarity 47.4%; Pred. No. 0.56;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 157 GCTAACGCTGCGCTTCTCTGCAAGCATGCCCCGTAATCTGAAGACCATTAACCAG 216
DB 4432 GTTACGGCTGCGCTGCTGTGAGCAGATGAGAGAGAGCTGAGAGGCTACAAATCCG 4491
QY 217 AGCGTTATGTTAAGCGCGCATGTAGCGCAGGTCGAGATATAGTACTATTAATG 276
DB 4492 ATTGCTCTCTANTTGGGAAGGATATGAGATGCTCGATTTTGTAGATGAGATG 4551
QY 277 ACTCAGATGTTTCAAGAAATATGCAATGCAATGCAATGCAATGCAATGCAATG 336
DB 4552 AAGAGGATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4611
QY 337 ATTACTGCGCGCAATGCGCGGTAATAGC 367
DB 4612 TTGCTGTGAAGATATGATGATGAGAAAG 4642

RESULT 6
US-09-220-132-30

; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
APPLICANT: Shyjan, Andrew W.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
FILE REFERENCE: 07334-074001
CURRENT APPLICATION NUMBER: US/09/220,132
CURRENT FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: US 60/079,303
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: US 60/068,821
PRIOR FILING DATE: 1997-12-24
NUMBER OF SEQ ID NOS: 191
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 14770
TYPE: DNA
ORGANISM: *Homo sapiens*
US-09-220-132-30

Query Match 7.3%; Score 33.4; DB 4; Length 14770;
Best Local Similarity 48.7%; Pred. No. 0.8;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 220 GGTATGTTAAGCGCGCATGTAGCGCAGGTCGAGATATAGTACTATTAATGACT 279
DB 10282 GGAATCTGGAAGAAATATGATTTGGCTTGGCTTGTACCTCAGAAATGAATGCC 10341
QY 280 CAGATGTTTCAAGAAATATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 339

Db 10342 CAGATGGGAGACAGACAGCAGTCCATCCAGAGTGTTCATTCGACACAGAGAGTTT 10401
QY 340 ACTGTCGGCAGATATACAGGAGGATATACAGCCGCGTGGTTATATATATGATCAGCTGTACC 399
Db 10402 TCTCATGACAGGATGACACAGATGATCTCTTGATGGCTATATGACCTGCAGATGATA 10461
QY 400 CGTGTG 406
Db 10462 GATGATG 10468

RESULT 7
US-09-489-039A-3211/C
Sequence 3211, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT FILING DATE: US/09/489,039A
PRIOR APPLICATION NUMBER: 2000-01-27
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 3211
LENGTH: 1122
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-3211

Query Match 7.1%; Score 32.2; DB 4; Length 1122;
Best Local Similarity 61.2%; Pred. No. 0.49;
Matches 52; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 311 ACCGATGGAACGCTAAAACTCCGATATTACTGTCGCGCAATACGCGCGTAAATACGCGG 370
Db 1007 AGCAGCGGACCGATTAACAGCGCGATCATGCCGAGGCGATGAGCGCGCAATAGCGCGG 948
QY 371 CGGTGTTATATATATGATCAGCTGT 395
Db 947 GTGAGATTAAGATCATTTGGCAGGT 923

RESULT 8
5486473-3/C
Patent No. 5486473
APPLICANT: FUKITA, HIROYUKI; YOSHIDA, IMO; TAKAGI, MITSUO;
MANABE, SADAO; FUKAI, KONOSUKE
TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
NUMBER OF SEQUENCES: 16
CURRENT APPLICATION DATA:
APPLICANT: Patterson, Robert, E.
FILING DATE: 09-FEB-1994
PRIOR APPLICATION NUMBER: US/08/194,049
FILING DATE: 18-DEC-1991
APPLICATION NUMBER: 809,255
FILING DATE: 05-DEC-1988
APPLICATION NUMBER: 932,419
FILING DATE: 19-NOV-1986
SEQ ID NO: 3
LENGTH: 1491
5486473-3

Query Match 6.8%; Score 31.2; DB 6; Length 1491;
Best Local Similarity 50.0%; Pred. No. 1.3;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
QY 178 CAAAGGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTATATGTAACGCGCC 237
Db 204 CAAATGACGACGAGGAGGATGTAACGATGCGGACATCTTGCAGAGGTGCTGCTTC 145

QY 238 GATGTAGGCGAGGATGCGGATATAGTACTATTAAGTACTGACATGAGATGTTTCAAGAAAT 297
Db 144 CATGTTCATCATTTTGAATCATGATGCTTGTCTTTTGAATATATATGATGATGATCACT 85
QY 298 AATGCCACATGACGACGATGAGCGTAAACTCC 333
Db 84 ATGCGCTTCAGTACAGATCAACCATCAATGATGCTCC 49

RESULT 9
US-09-540-236-492/C
Sequence 492, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI
FILE REFERENCE: 2709.2005-001
CURRENT FILING DATE: US/09/540,236
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 492
LENGTH: 2436
TYPE: DNA
ORGANISM: M. catarrhalis
US-09-540-236-492

Query Match 6.8%; Score 31.2; DB 4; Length 2436;
Best Local Similarity 53.7%; Pred. No. 1.7;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;
QY 209 TTACCGAGGCGTTATGCTTAAGCGCGCCGATGAGCCAGGATGCGGATATATGACTA 268
Db 1511 TTACCGATTCATCATTAATCTTAAGGCAAGTGGGATGCTGATGATTAATGACG 1452
QY 269 TTGA---ACTGACTCAATGTTTCAGAAATTAATGCCACATGACGATGAGGAAAGCTA 325
Db 1451 TGAATTCATTAATCAAAATGCTTCTATTAATTAAGCCAGTATCCCTCTGACAGAA 1392
QY 326 AAAATCCGATTAATGCTGCGCAATACGCGGATTAATAGGCC 369
Db 1391 ACTGTACCTCATGACATCAACCAAACTTCGTTGATTC 1348

RESULT 10
US-09-596-002-31/C
Sequence 31, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Patterson, Robert, E.
FILING DATE: 09-008-4 US
FILE REFERENCE: PM-008-4 US
CURRENT FILING DATE: US/09/596,002
PRIOR APPLICATION NUMBER: 2000-06-16
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 31
LENGTH: 65792
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte template ID No. 6632636 31
PUBLICATION INFORMATION:
US-09-596-002-31

Query Match 6.8%; Score 31.2; DB 4; Length 65792;
Best Local Similarity 53.7%; Pred. No. 11;

```

Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 280 CAGAAATGTTTCAGAAATATATGCAACCATCGACAGTGGAAACGCTAAAACTCCGATATT 339
Db 876 CAGATGGTTGACCAATTTGAGACCAATGATGACGAGTACGCGCATTAAGAACGCCATT 935
QY 340 ACTGTCGGCCATACGGCGCGTAAATPAGCC 369
Db 936 GTTGTGCTCCCAATCGGTGGCGAAGACTCC 965

RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R. M.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 6.7%; Score 30.6; DB 3; Length 4403765;
Best Local Similarity 51.9%; Pred. No. 95;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 45 TGGCAGTGTCTTGCGCTGCGCTGTTCCACAAATGGGGCGCGCGTAAATCAACGCGCG 104
Db 1357218 TTGCGGAGCGGTGCTGCTACGCGCGCGCTCACACGCGGGGCGTGGTACGAGTCCGTCGA 13571598
QY 105 CGCAATATGTTCCGGCCCGGACTCAAGTTCAGCATTTATCAGTAGCGATTCCGCTAACGC 164
Db 1357158 AAACACCGCATGCGGTACGACGCCAGATTAAGTGTGGCCATGTCTATTTGTTGGCGC 13570999
QY 165 TGCGCTTGCTCTG 177
Db 1357098 CCGGAGCTGCTG 1357086

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-2007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA

```

```

; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

Query Match	6.7%;	Score 30.6;	DB 3;	Length 441529;
Best Local Similarity	51.9%;	Pred. NO. 95;		
Matches 69;	Conservative 0;	Mismatches 64;	Indels 0;	Gaps 0;

OY 45 TTGCGAGTGTCTGTGGCTGGCGCTGTTCACAAATGGGCGCGCGCGGTATTCATTAACGGCGG 104
 Db 1357750 TTGCGGAGCGCTGCTGACGCGCGCGCGCTACACGCGCGCGCTGTGTGTACGAGTCCGTGA 1357691

105 CGGAAATAGTTCGGCCCGGACATCAACCTGTACCATTTACGTACCGTTCGCTAACGC 164
 D5 1357680 AAACACCGCATGCGTAAACGAGCCACCATTAAGTGTGGCCATGTCGATCTGTGTGGCC 1357691

QY 165 TCGGCTTGCCTCTG 177
|||
Db 1357630 CCCGGACGTGCTG 1357618

RESULT 14
US-09-350

```

Sequence 2 Application US/09350756
Patent No. 6495143
GENERAL INFORMATION:
APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
APPLICANT: John S. Lee
APPLICANT: Peter Pushko
APPLICANT: Michael D. Parker
APPLICANT: Jonathan F. Smith
APPLICANT: Mark T. Dertzbaugh
APPLICANT: Leonard Smith
TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
FILE REFERENCE: 003/124/SAP RIID 98-21
CURRENT APPLICATION NUMBER: US/09/350,756
CURRENT FILING DATE: 1999-07-09
EARLIER APPLICATION NUMBER: US 60/092,416
EARLIER FILING DATE: 1998-07-10
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Apple Macintosh Microsoft Word 6.0
SEQ ID NO 2
LENGTH: 1341
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-2

```

Query Match	6.6%	Score 30;	DB 4;	Length 1341;
Best Local Similarity	53.4%;	Pred. No. 3.2;		
Matches 63;	Conservative 0;	Mismatches 55;	Indels 0;	Gaps 0;

y 330 CTCGGATATTACTGTCGCCAATAAGCGCGGTATTAATGATCA 38
 |||||
 Db 673 CTTGCATTTGGACTGAGACAGTCCGCTTGGAAGATGAGAATATTCATCCAGATGA 614

Dy 390 GCTGGTACCCGCTGTGTATTACCATGAATGGCAGTCAGAACAGCCCGCTTAAC 447
|||||
Db 613 ACTGGTAGCATGCATGTCACCCGTCAGTTTGAGATGATTTCCACCGTTACCGATTAAC 556

RESULT 15
US-09-252

US-09-252-991A,10904
Sequence 10904, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196,136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18

```

? PRIOR APPLICATION NUMBER: US 60/074,768
? PRIOR FILING DATE: 1996-02-18
? PRIOR APPLICATION NUMBER: US 60/094,100
? PRIOR FILING DATE: 1996-07-27
? NUMBER OF SEQ ID NOS: 33142
? SEQ ID NO 10904
? LENGTH: 1008
? TYPE: DNA
? ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

```

Query Match	6.5%	Score 29.8;	DB 4;	Length 1008;
Best Local Similarity	54.0%;	Pred. No. 3.2;		
Matches 61; Conservative	0;	Mismatches 52;	Indels 0;	Gaps 0;

0y 41 TTTCTGGCAGTCTCTGGCGCGCTTCACAAATGGGCGCGCGGGTATCATTAAC 100
Db 141 TGTATGCCCTGCTCGGTGACCCGAGCGCGCTTCGCGCGCGAAGCCGCTGG 200

```

0y      101  GCGGCGGCATTAGTTCGGGCCCGCACTCAACGTTGAGCATTATCAAGTACGGT 153
          |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||
Db      201  GCGGGGAGGAGAGCCCGCACCCTGTCGCGCGCTTGCGCTTGATCAAGCAGAT 253

```

Search completed: March 16, 2004, 04:37:19
Job time : 63.6647 secs

QY 52 GCTCGAGCTGGCGTGGTCTTCACAAATGGGGCGCGCGGTATCATATACGCGCGGCNAAT 111
 DB 209 GCTCGAGTTCGCCGATTAATCTGATTTGGCGCATATGACAGTGAATTATTAACAGTTCCGCCNAAT 156
 QY 112 AGTTCCGCGCCCGAGCTCAACGTTGAGCACTTTATCACTAGCGTTCCGCTTAACGCTGCGCTT 171
 DB 149 AATTGAGCCATTGCTCAATGTTTGGCGATCGATGGTGTGGCCCAACGAGAAATTT 90
 QY 172 GC 173
 DB 89 GC 88
 RESULT 4
 US-09-879-248-5/c
 / Sequence 5, Application US/09879248
 / Patent No. US20020062530A1
 / GENERAL INFORMATION:
 / APPLICANT: Fan, Hao
 / APPLICANT: Wei, Zhong-Min
 / TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
 / FILE REFERENCE: 21829/81
 / CURRENT APPLICATION NUMBER: US/09/879,248
 / PRIOR FILING DATE: 2001-06-12
 / PRIOR APPLICATION NUMBER: 60/212,211
 / NUMBER OF SEQ ID NOS: 16
 / SOFTWARE: PatentIn Ver. 2.1
 / SEQ ID NO 5
 / LENGTH: 1344
 / TYPE: DNA
 / ORGANISM: Erwinia amylovora
 US-09-879-248-5
 Query Match 7.5%; Score 34; DB 9; Length 1344;
 Best Local Similarity 54.9%; Pct. No. 0.93;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;
 QY 52 GCTCGAGCTGGCGTGGTCTTCACAAATGGGGCGCGCGGTATCATATACGCGCGGCNAAT 111
 DB 209 GCTCGAGTTCGCCGATTAATCTGATTTGGCGCATATGACAGTGAATTATTAACAGTTCCGCCNAAT 156
 QY 112 AGTTCCGCGCCCGAGCTCAACGTTGAGCACTTTATCACTAGCGTTCCGCTTAACGCTGCGCTT 171
 DB 149 AATTGAGCCATTGCTCAATGTTTGGCGATCGATGGTGTGGCCCAACGAGAAATTT 90
 QY 172 GC 173
 DB 89 GC 88
 RESULT 5
 US-10-010-390-6/c
 / Sequence 6, Application US/10010390
 / Patent No. US20030104979A1
 / GENERAL INFORMATION:
 / APPLICANT: Wei, Zhong-Min
 / APPLICANT: Leon, Ernesto
 / TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
 / FILE REFERENCE: 21829/111
 / CURRENT APPLICATION NUMBER: US/10/010,390
 / PRIOR FILING DATE: 2001-11-05
 / PRIOR APPLICATION NUMBER: 60/248,169
 / NUMBER OF SEQ ID NOS: 14
 / SOFTWARE: PatentIn Ver. 2.1
 / SEQ ID NO 6
 / LENGTH: 1344
 / TYPE: DNA
 / ORGANISM: Erwinia amylovora

```

QY      238 GATGTAGGCCAGGGTGCGGATAATAGTACTATTGAACTGACTCAGAATGGTTTCAGAAAT 297
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB      321 GATGTAGTGAAGAAGCGAAAAATTACTATCGACAAACGGTTTCTCGTATGACATAT 380

```


Query Match 7.3%; Score 33.4; DB 12; Length 7766;
Best Local Similarity 47.4%; Pred. No. 4;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

157 GCTAACGCTGGCTGCTCTGCAAGCGATCCCGTAATCTGAAGACACCATTCACG 216
4432 GTTACGGCTGGCGCTGCTGAGACAGATGAGAGAAAGCTGCAAGGCTACCAATCCG 4491
217 AGCGTTATGTTAAGCGCGCGATGTAGCCAGGCTGGATATATAGTACTATTGAACG 276
4492 ATTCTGCTGCTATTTGGAGAGGCTATGAGATGCTGGGATTTTGTGAAGATGAGATG 4551
277 ACTAGAGATGTTTCAGAAATATATGCAACCATGCAACGCTGAAGCTGAAGCTCCGAT 336
4552 AAGAAAGATGATCAAGATGCTGCTGCTATTCCTTTGAGGGGATGTGCTAAGGATGGAAG 4611
337 ATTACTGTGGCCCAATACGCGCGTAAATACG 367
4612 TTGCTGTGAAGAGTAAATGATGTGAGAAAG 4642

RESULT 12

US-10-222-162-3
Sequence 3, Application US/10222162
Publication No. US20030060618A1
GENERAL INFORMATION:
APPLICANT: NORRIS, STEVEN J.
APPLICANT: JING-REN, ZHANG
APPLICANT: HARDHAM, JOHN M.
APPLICANT: HOWELL, JERRILYN K.
APPLICANT: BARBOUR, ALAN G.
APPLICANT: WEINSTOCK, GEORGE M.
TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
FILE REFERENCE: UTS# 234USD4
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: 09/125,619
PRIOR FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 50
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 7766
TYPE: DNA
ORGANISM: Borrelia burgdorferi
FEATURE:
NAME/KEY: modified_base
LOCATION: (127)
OTHER INFORMATION: R = A OR G
US-10-222-162-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 4;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

157 GCTAACGCTGGCTGCTCTGCAAGCGATCCCGTAATCTGAAGACACCATTCACG 216
4432 GTTACGGCTGGCGCTGCTGAGACAGATGAGAGAAAGCTGCAAGGCTACCAATCCG 4491
217 AGCGTTATGTTAAGCGCGCGATGTAGCCAGGCTGGATATATAGTACTATTGAACG 276
4492 ATTCTGCTGCTATTTGGAGAGGCTATGAGATGCTGGGATTTTGTGAAGATGAGATG 4551
277 ACTAGAGATGTTTCAGAAATATATGCAACCATGCAACGCTGAAGCTGAAGCTCCGAT 336
4552 AAGAAAGATGATCAAGATGCTGCTGCTATTCCTTTGAGGGGATGTGCTAAGGATGGAAG 4611
337 ATTACTGTGGCCCAATACGCGCGTAAATACG 367
4612 TTGCTGTGAAGAGTAAATGATGTGAGAAAG 4642

RESULT 13

US-10-143-024-3

Sequence 3, Application US/10143024
Publication No. US20030092903A1
GENERAL INFORMATION:
APPLICANT: NORRIS, STEVEN J.
APPLICANT: JING-REN, ZHANG
APPLICANT: HARDHAM, JOHN M.
APPLICANT: HOWELL, JERRILYN K.
APPLICANT: BARBOUR, ALAN G.
APPLICANT: WEINSTOCK, GEORGE M.
TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
FILE REFERENCE: UTS# 234USD1
CURRENT FILING DATE: 2002-05-10
PRIOR APPLICATION NUMBER: 09/125,619
PRIOR FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 7766
TYPE: DNA
ORGANISM: Borrelia burgdorferi
FEATURE:
NAME/KEY: modified_base
LOCATION: (127)
OTHER INFORMATION: R = A OR G
US-10-143-024-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 4;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

157 GCTAACGCTGGCTGCTCTGCAAGCGATCCCGTAATCTGAAGACACCATTCACG 216
4432 GTTACGGCTGGCGCTGCTGAGACAGATGAGAGAAAGCTGCAAGGCTACCAATCCG 4491
217 AGCGTTATGTTAAGCGCGCGATGTAGCCAGGCTGGATATATAGTACTATTGAACG 276
4492 ATTCTGCTGCTATTTGGAGAGGCTATGAGATGCTGGGATTTTGTGAAGATGAGATG 4551
277 ACTAGAGATGTTTCAGAAATATATGCAACCATGCAACGCTGAAGCTGAAGCTCCGAT 336
4552 AAGAAAGATGATCAAGATGCTGCTGCTATTCCTTTGAGGGGATGTGCTAAGGATGGAAG 4611
337 ATTACTGTGGCCCAATACGCGCGTAAATACG 367
4612 TTGCTGTGAAGAGTAAATGATGTGAGAAAG 4642

RESULT 14

US-09-873-367C-230
Sequence 230, Application US/09873367C
Publication No. US20030165839A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
APPLICANT: Soppet, Daniel
APPLICANT: Endress, Gregory
APPLICANT: Augustus, Meena
APPLICANT: Ebner, Reinhard
APPLICANT: Carter, Kenneth
TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
FILE REFERENCE: 689290-64
CURRENT FILING DATE: US/09/873,367C
PRIOR APPLICATION NUMBER: 2003-04-29
PRIOR FILING DATE: 2000-09-29
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/236,842
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: U.S. 60/244,867
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: U.S. 60/245,084
PRIOR FILING DATE: 2000-11-01

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgc.....ccacgcctaacagatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrc:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rod:*
26: em_ges_phg:*
27: em_ges_vr1:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216.6	47.5	680	12	BU618688 BU618688
2	35.2	7.7	813	28	B2204853 CH230-390
3	34.8	7.6	801	12	B1648266 603278209
4	34.2	7.5	523	12	BU334624 BU334624

5	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
6	34	7.5	234	10	BS581987	BS581987
7	34	7.5	500	14	CA711477	CA711477 wdk2c.pk0
8	34	7.5	811	28	B2469000	B2469000 BOAO58TF
9	33.8	7.4	296	9	AV058630	AV058630
10	33.8	7.4	365	9	AA746477	AA746477 nwe2e02.8
11	33.8	7.4	558	14	CD374421	CD374421 TMBMFC2H
12	33.8	7.4	559	13	BU765156	BU765156 8a823f11.
13	33.8	7.4	695	28	AQ637195	AQ637195 RPCT-11-4
14	33.8	7.4	941	29	CG178280	CG178280 PUUDF79TD
15	33.8	7.4	1098	29	CNS02GQP	AL196666 Tetracodon
16	33.6	7.4	374	12	BY236219	BY236219 BY236219
17	33.6	7.4	594	13	BJ335653	BJ335653 BJJ35653
18	33.6	7.4	653	12	BJ336964	BJ336964 BJJ36964
19	33.6	7.4	712	14	CF437982	CF437982 EST674327
20	33.4	7.3	505	6	AL809250	AL809250 TITicium
21	33.4	7.3	624	14	CB578319	CB578319 AMGNNUC:N
22	33.4	7.3	664	13	BX506928	BX506928 DKFZP7791
23	33.4	7.3	679	29	CC961526	CC961526 BOTG510TF
24	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
25	33.4	7.3	699	12	B1959328	B1959328 HVSMEN001
26	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
27	33.4	7.3	709	28	BH690914	BH690914 BOHXR64TF
28	33.4	7.3	715	13	BU444683	BU444683 603765715
29	33.4	7.3	785	28	BH578793	BH578793 BOGLV20TR
30	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
31	33.4	7.3	827	28	BH438448	BH438448 BOGVQ28TR
32	33.4	7.3	830	28	BH715216	BH715216 BOIMV42TR
33	33.4	7.3	835	28	BH575194	BH575194 BOHJ027TF
34	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
35	33.4	7.3	856	14	CD456845	CD456845 FG05_03n1
36	33.4	7.3	1026	28	BZ458784	BZ458784 BOMHB14TF
37	33.2	7.3	589	28	BZ300068	BZ300068 KO0510.P1
38	33.2	7.3	592	29	CG072975	CG072975 PUCM43TB
39	33.2	7.3	629	9	AU295325	AU295325 AU295325
40	33.2	7.3	675	28	AQ361497	AQ361497 mxb00004K
41	33.2	7.3	788	9	AU139321	AU139321 AU139321
42	33.2	7.3	1646	10	BF579120	BF579120 602096054
43	33	7.2	456	12	BU331339	BU331339 BU331339
44	33	7.2	524	12	BU367372	BU367372 BU367372
45	33	7.2	547	12	BU333958	BU333958 BU333958

ALIGNMENTS

RESULT 1
LOCUS BU618688/c 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BU618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186D22 5', mRNA sequence.
ACCESSION BU618688
VERSION BU618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis

REFERENCE
AUTHORS Kohata,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
CONTACT Tadao Shin-1
CENTER For Genetic Resource Information
NATIONAL Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
TEL: 81-559-81-6856
FAX: 81-559-81-6855
EMAIL: tshin1@genie.nig.ac.jp
The information of this clone is available through the following URL.

http://xenopus.nibb.ac.jp.

FEATURES

Location/Qualifiers

1. 680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="X186B22"
/issue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_1db="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match

Best Local Similarity 47.5%; Score 216.6; DB 12; Length 680;
Matches 306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTAAAGTGGAGCATTGCGACATGTTGTTGCGAGTCTGGCT 60
DB 593 ATGAACTTTAAAGTGGAGCATTGCGACATGTTGTTGCGAGTCTGGCT 60
QY 61 GGCGTCTTCCACATGGGGCGGGCGGTAATCAATACGGCGGCGCAATAGTCCGCG 534
DB 533 GGCGTCTTCCACATGGGGCGGGCGGTAATCAATACGGCGGCGCAATAGTCCGCG 534
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGGCGGTAATAGCGCG 474
DB 473 CCAAACTTGTAGTGACATTTTACAGTACGTTCCGCTAACGCTGGCGGTAATAGCGCG 474
QY 181 AGCGATGCCCGTAAATCTGAAACGACATTACCGAGCGGTTATGTTACGGCGCGAT 414
DB 413 AGCGATGCCCGTAAATCTGAAACGACATTACCGAGCGGTTATGTTACGGCGCGAT 414
QY 241 GTAGGCGAGGCGTCCGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 354
DB 353 GTAGGCGAGGCGTCCGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 354
QY 301 GCCACCATGACGCGTGGTAAAGCTGAAAGCTGAAAGCTGAAAGCTGAAAGCTGAAAGCT 294
DB 293 GCTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AATAAGCCCGCGTGTATATATGATGATGATGATGATGATGATGATGATGATGATGAT 234
DB 233 GCGAACGCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GCAATGCAACCAACGCGCGTAAACGATGATGATGATGATGATGATGATGATGATGAT 174
DB 173 GGCTTGTATGACACCGCGCTCATCATGATGATGATGATGATGATGATGATGATGAT 138

```

RESULT 2

BZ204853

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BZ204853 813 bp DNA linear GSS 11-OCT-2002
CH230-390D21, TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
BZ204853
BZ204853.1 GI:23862905
GSS.
Rattus norvegicus (Norway rat)
Rattus norvegicus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 813)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shivatsbeyn, A., Gebregiorgis, E., Overton, L., Russell, D., Chen, D.,
R1998, F., de Jong, P., and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other GSSs: CH230-390D21, TV.
Contact: Shanying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research

FEATURES

source

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cno.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or erting_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Seq primer: 77
Class: BAC ends.

ORIGIN

Query Match

Best Local Similarity 7.7%; Score 35.2; DB 28; Length 813;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```

QY 178 CAAGCCAGTCCCGTAAATGTAAGCAACATTAACCGAGCGGTATGTAACGGCGCC 237
DB 548 CAAGCCAGTCCCGTAAATGTAAGCAACATTAACCGAGCGGTATGTAACGGCGCC 237
QY 238 GATGTAGCCGAGGTCGCGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 297
DB 608 TCTGGAACACATGGAGGACGCGGAAAAAATCTCTGAACAAACCAATGGCTTATGCT 667
QY 298 AATGCCACATGACGACGATGTAAGCAACGTAAGCAACGTAAGCAACGTAAGCAACG 667
DB 668 AATGCCACATGACGACGATGTAAGCAACGTAAGCAACGTAAGCAACGTAAGCAACG 357
QY 358 GGTATTAACGCC 369
DB 728 TGTGTTAGGAC 739

```

RESULT 3

B1648266

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

B1648266 801 bp mRNA linear EST 12-SEP-2001
603278209P1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5318307 5',
mRNA sequence.
B1648266
B1648266.1 GI:15562502
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 801)
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-research.nih.gov
Tissue Procurement: Jochim Hennighausen Ph.D., Chu-Xia Deng Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/BLND at:

http://image.llnl.gov
 Plate: LHAM1807 row: b column: 04
 High quality sequence stop: 798.
 Location/Qualifiers

FEATURES

Source

```
1..801
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129.C57BL/6J.FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5318307"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-Sport6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."
```

ORIGIN

```
Query Match      7.6%; Score 34.8; DB 12; Length 801;
Best Local Similarity 50.0%; Pred. No. 76;
Matches 114; Conservative 0; Mismatches 112; Indels 2; Gaps 1;

OY 210 TACCAGAGCGGTTATGTTACGCGCGGATGAGCGGATGATTAATGACTAT 269
    |||||
DB 322 TTCACCCGCAAGATGTGACCGCTTATCCCTCGGCGCGGCACTGACCG 381
    |||||
OY 270 TGAATGACGATGATGTTTCAGAA--ATAATGCCACATGACCACTGGACGCTTAA 327
    |||||
DB 382 TGAATCTTCACAGATCATGACATGAGAAAGTCCCAAGACCGCAAGCCAG 441
    |||||
OY 328 AACTCCGATTAATGTCGCGCAATACGCGCGGTTAATTAATGAT 387
    |||||
DB 442 GACGCCACCTCGCAATAGCAAGACGCGCTGACCTCGCACTGCAAGATGAG 501
    |||||
OY 388 CAGCTGTTACCCGTTGTTACCATGAATGGACATGCAACAC 435
    |||||
DB 502 CTGCTGGTCCGCGATGAGAGGTTGAGACCAAGACGAGAGAC 549
    |||||
```

RESULT 4

LOCUS

Bj334624 523 bp mRNA linear EST 05-MAR-2002
 Bj334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
 discoideum cDNA clone dda47006 5', mRNA sequence.

ACCESSION

Bj334624 GI:19164754

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

Unpublished (2002)
 Contact: Tadabu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yatai, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers

FEATURES

Source

```
1..523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47006"
```

```
/sex="mat A"
/dev_stage="aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"
```

ORIGIN

```
Query Match      7.5%; Score 34.2; DB 12; Length 523;
Best Local Similarity 57.1%; Pred. No. 93;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 238 GATGAGCCAGGTTGGGATTAATGATCTTGAAGTCAAGATGTTTCAAAAT 297
    |||||
DB 398 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 457
    |||||
OY 298 AATGCCACATCGACCGATGAGCTTAAATCCGATATTAAT 342
    |||||
DB 458 AATAACACACACACACATTAATTAATTAATTAATTAATTAATTAATTAAT 502
    |||||
```

RESULT 5

LOCUS

BG671547 1036 bp mRNA linear EST 30-APR-2001
 DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5',
 mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

COMMENT

1 (bases 1 to 1036)
 Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C.,
 Yang,L., Huang,W.J., Pu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D.,
 Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
 Identification of gene expression profile of dorsal root ganglion
 in the rat peripheral axotomy model of neuropathic pain
 Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
 22056133
 JOURNAL MEDLINE
 PUBMED
 12060780
 Contact: Zhang Xu
 Laboratory of Sensory System
 Institute of Neuroscience
 320 Yue Yang Road, Shanghai 200031, P.R.China
 Tel: 86-21-64748700-121
 Fax: 86-21-64713446
 Email: xu.zhang@ion.ac.cn
 This clone is also available at Chinese National Human Genome
 Center at Shanghai, 351 Guo Shoujing Road, Zhongjiang Hi-Tech Park,
 Pudong New Area, P.R.China. Please contact with Zhang Xu
 (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
 PCR Primers
 FORWARD: T3
 BACKWARD: T7
 Seq primer: T3
 POLYA=NO.

FEATURES

Source

```
1..1036
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRNBUB05"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/clone_lib="Rat DRG Library"
```

ORIGIN

```
Query Match      7.5%; Score 34.2; DB 12; Length 1036;
Best Local Similarity 49.2%; Pred. No. 1.3e+02;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
```

QY 11 TAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGAGCGCTCTGCGCTCGTTC 70
 |||||
 Db 168 TGAACATACATGAGTGAAGTGTCCGATTCGACAGATTCGTATGATGCGTCTGT 227
 |||||
 QY 71 CACAAATGGGCGCGCGGATATCATATACGGGCGGGAATATGTTCCGCCGACTCAA 130
 |||||
 Db 228 CTGTAGAGTCTCTCCCTGGTAAAGTCCGAGATGCATCGATATGCTGGCGGTTA 287
 |||||
 QY 131 CGTTAGCATTTATAGTACGCTTCGCTTAACGCTGCGCTTCTCTGCAAGGATGCC 190
 |||||
 Db 288 CATCGAAGTGAAGAGGTGGCTCTTTCGAACTCTGACTTGTATCTGTCAAGGAGCGTT 347
 |||||
 QY 191 GTA 193
 |||||
 Db 348 GTA 350

RESULT 6
 BB581987/c 234 bp mRNA linear EST 30-NOV-2000
 LOCUS BB581987 RIKEN full-length enriched, adult male colon Mus musculus
 DEFINITION CDNA clone 9030005005 5', mRNA sequence.
 ACCESSION BB581987
 VERSION BB581987.1 GI:11478531
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 234)
 Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T.,
 Carninci, P., Hanagaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T.,
 Hodojima, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, J.,
 Kojima, Y., Komoto, H., Kusakabe, M., Matsuyama, T., Miyazaki, A.,
 Okamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, Y.,
 Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K.,
 Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,
 Suzuki, H., Tagawa, A., Tanahashi, F., Tanaka, T., Toya, T.,
 Watanabe, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A.,
 Muramatsu, M., and Hayashizaki, Y.
 RIKEN Mouse ESTs (Aizawa, K. et al. 2000)
 Unpublished (2000)
 COMMENT Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic
 Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@gsr.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/
 Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 Thermostabilization and thermoactivation of thermostable enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
 Okazaki, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci, P. and Hayashizaki, Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.

FEATURES

source
 1..234
 location/Qualifiers
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="9030005005"

ORIGIN

Query Match 7.5%; Score 34; DB 10; Length 234;
 Best Local Similarity 51.3%; Pred. No. 73;
 Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

QY 35 TCGTACTTTCTGCGAGTGTCTGGCTGCGCTGTCACAAATGGGCGCGCGGTATC 94
 |||||
 Db 200 TAGTTTGGTAGTGTGCGGTTATGCGGTTCTCTTCCACACAGCCGCCGCTTTC 141
 |||||
 QY 95 ATACGCGCGCGCGCATATGTTCCGCCCGGAGCTAAGCTTATGATGATGATC 154
 |||||
 Db 140 CGAAGCGCGGCGCTTTTATGCTTTGTTCAAAATATGATGCGCATTTACCACTTT 81
 |||||
 QY 155 CGCTAAGCGTGGCGCTTCTGTCGAAGCGATGC 188
 |||||
 Db 80 CCGTTACCAAGCCCGTGGCCGCGCATTTGGCGGC 47
 |||||

RESULT 7

CA711477/c 500 bp mRNA linear EST 26-NOV-2002
 LOCUS CA711477
 DEFINITION wdk2c.pk014.o3 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.o3 5'
 end, mRNA sequence.
 ACCESSION CA711477
 VERSION CA711477.1 GI:25433270
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 1 (bases 1 to 500)
 Tingley, S. V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
 Miao, G., Caraher, N., and Hanafey, M. K.
 Dupont Wheat cDNA Sequence
 Unpublished (2002)
 COMMENT Contact: Scott V. Tingley
 Crop Genetics
 E. I. Dupont de Nemours and Company
 1 Innovation Way P.O. Box 6104, Newark, DE 19714-6104, USA
 Tel: 302-631-2602
 Fax: 302-631-2607
 Email: Scott.V.Tingley@USA.dupont.com
 Seq primer: M13.
 location/Qualifiers
 1..500
 Location/Qualifiers

FEATURES

source
 1..500
 location/Qualifiers
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"

/clone="wtk2c.pk014.o3"
/tissue type="kernel"
/clone lib="wtk2c"
/note="Vector: pBluescript SK+, Site 1: EcoRI, Site 2:
XhoI; wheat (Triticum aestivum L.) developing kernel, 7
days after anthesis."

ORIGIN

Query Match 7.5%; Score 34; DB 14; Length 500;
Best Local Similarity 54.5%; Pred. No. 1e+02;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 15 AGTGGACGATTCGACGATCTGAGTCTGCTGGCTGCTGTTCCACA 74
Db 274 AGTGGATGACATCCGACGCGGCTCTCCCGCGCTTTCGAACCGCGGAGCTCCCN 215
Qy 75 ATGGGGCGCGCGGATATCAACGCGCGGCAATGTTCCGCGCGAC 126
Db 214 TGTGGAGGCGGTGGGATGACATGCGCGCGCGCGGTGGCGCCAC 163

RESULT 8
B246900/c 811 bp DNA linear GSS 13-DEC-2002
LOCUS BOOAO58TF.BQ.1.6.2.KB.tot.Brassica oleracea genomic clone BOOAO58,
DEFINITION genomic survey sequence.
ACCESSION B246900
VERSION B246900.1 GI:26764546
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 811)
TOWN,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSSes: BOOAO58TR
Contact: Chris Town

REFERENCE 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tol: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES

SOURCE

Location/Qualifiers
1..811
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOOAO58"
/clone_1lb="BO.1.6.2.KB.tot"
/note="Vector: pHOSt; Site 1: BstXI, 1.6-2 kb sheared
total DNA inserted into pHOSt using BstXI linkers"

ORIGIN

Query Match 7.5%; Score 34; DB 28; Length 811;
Best Local Similarity 48.9%; Pred. No. 1.3e+02;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 176 TCGAAGCGATCCGTAATCTGAACGACCATTAACGAGCGGTTATGTAACGCG 235
Db 610 TGTAGGCAAGCATGTAAACTGTGTTCCCATTCATCACTGTTATGACAAATGCT 551
Qy 236 CCGATGAGGCGGCGGATATAGTACTATTTGAACGACCTCAAGATGTTTCAGAA 295
Db 550 TTGATCTTAATTCATCTGATGTTTGACTGCTCTGCTGTTAGGATATGTTTAAT 491
Qy 296 ATATGCGACCATGACGACGATGAACGTAATAAATCCGATATTAATGCGGCAATACG 355

Db 490 ACTTTGCACATTCATTAATGAATAATCGAAATACCTGATACACTTATTCAAACA 431
Qy 356 GCGGTA 361
Db 430 AAGATA 425

RESULT 9
AV058630 296 bp mRNA linear EST 23-JUN-1999
LOCUS AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
DEFINITION clone 1810054H02, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 296)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
Hara,A., Hayatsu,N., Hori,F., Iehikawa,T., Itoh,M., Izawa,M.,
Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H.,
Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
Suganara,Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y.,
Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-resortc.riken.go.jp
Thermolabile and thermostable activation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES

SOURCE

Location/Qualifiers
1..296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue type="pancreas"
/dev stage="adult"
/clone_1lb="Mus musculus pancreas C57BL/6J adult"

ORIGIN

Query Match 7.4%; Score 33.8; DB 9; Length 296;
Best Local Similarity 53.4%; Pred. No. 93;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 251 GTGGGATATATGATCTAGTAACTGACTGAGATGTTTCAGAAATATGACCACTCG 310
Db 13 GTGAGACCAAGATTTTCTTAGGTACCAAGGCTGGTAAAGAAAGACATCAACTG 72
Qy 311 ACAGTGAACGCTAAATACTCGATATTAATCTGCGGCAATACGCGGTAATTAACGCG 370
Db 73 TCCAGTGAACCAAAATAAGAAAGAAATATTAATCAAAAGTTGAATTAAGTAG 132
Qy 371 CGCTGTTAATTA 383
Db 133 AGATGCCAAATA 145

LOCUS	CD374421/c	558 bp	mRNA	linear	EST 19-AUG-2003
DEFINITION	TNMBMFC2H33 Wuchereria bancrofti microfilaria cDNA (SAW55JL-WBMF)				
ACCESSION	TNMBMFC2H33 Wuchereria bancrofti cDNA clone TNMBMFC2H3 5', mRNA sequence.				
VERSION	CD374421				
KEYWORDS	CD374421.1 GI:31229939				
SOURCE	EST.				
ORGANISM	Wuchereria bancrofti				
REFERENCE	Wuchereria bancrofti				
AUTHORS	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;				
TITLE	Onchocercidae; Wuchereria.				
JOURNAL	1 (bases 1 to 558)				
COMMENT	Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.				
	Toward the characterization of the Wuchereria bancrofti				
	microfilarial transcription plasmids with comparisons to those of Brugia				
	malayi and Onchocerca volvulus				
	Unpublished (2003)				
	Contact: Thomas B. Nutman				
	Laboratory of Parasitic Diseases				
	NII				
	Building 4 Room 126, Bethesda, MD 20892-0425, USA				
	Email: tnutman@niaid.nih.gov				
	Seq primer: Pbluescript T3.				
FEATURES	Location/Qualifiers				
source	1..558				
	/organism="Wuchereria bancrofti"				
	/mol_type="mRNA"				
	/db_xref="taxon:6293"				
	/clone="TNMBMFC2H3"				
	/dev_stage="microfilaria"				
	/lab_host="XLI-Blue MRF"				
	/clone_id="Wuchereria bancrofti microfilaria cDNA				
	(SAW55JL-WBMF)"				
	/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:				
	Xho I; Lymphatic filarial nematode parasite of humans				
	mRNA was prepared from approximately 85,000 microfilariae				
	isolated from the blood of an infected individual from				
	Guyana, South America and converted to double-stranded				
	cDNA using reverse transcriptase and oligo(dT) followed by				
	Rnase H and DNA pol I. The library has 2.9 x 10E5				
	independent recombinants and the average insert size is				
	~1kb. The library was constructed by Sandra J. Laney. The				
	library is available from Dr.S.A.Williams, email:				
	genome@smith.edu."				
ORIGIN					
Query Match	7.4%; Score 33.8; DB 14; Length 558;				
Best Local Similarity	60.7%; Pred. No. 1.2e+02;				
Matches 65; Conservative	0; Mismatches 40; Indels 2; Gaps 1;				
Oy	76 TGGGGGGGGGGCGGTATCATTAACGGCGCGGCAATAGTCCGGCCCGGACTCAACGTTG 135				
Db	161 TGGGGGGGGGGCGGTATCAACCGCGGTGGCGGTAATTAATTAACGGCGCCCAATTCATGCT 102				
Oy	136 AGCAATTATC--AGTACGGTTCGCGTAACGCTGCGGCTGCTCGTGCAT 180				
Db	101 GACATTATNCCAAGANCGGTGGCGGGANATCTGCNCTTCTCTGNA 55				
RESULT 12					
LOCUS	BU765156 559 bp mRNA linear EST 10-OCT-2002				
DEFINITION	saas2f11.y1 Gm-cl080 Glycine max cDNA clone SOYBEAN CLONE ID:				
	Gm-cl080-4669 5' similar to SW:DI00.ARAH 000874				
	DNA-DAMAGE-REPAIR/TOLERATION PROTEIN DRT100 PRECURSOR ;, mRNA				

rosids; eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

1 (bases 1 to 559)

Shoemaker, R., Keim, P., Vockin, L., Erpelting, J., Coryell, V., Klamma, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Persen, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Riter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R., and Wilson, R.

Public Soybean EST Project

Unpublished (1999)

Contact: Shoemaker R/Public Soybean EST Project

Public Soybean EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available through: RegGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@reggen.com web site: www.reggen.com

Seq primer: -40RP from Glibco

High quality sequence stop: 448.

Location/Qualifiers

1. 559

/organism="Glycine max"

/mol_type="mRNA"

/db_xref="taxon:3847"

/clone="SOYBEAN CLONE ID: Gm-c1080-4869"

/tissue_type="Roots of a 8 day old 'Brags' supermodulating mutant NTS382 seedlings"

/dev_stage="8 days old"

/lab_host="DH10B"

/clone_id="Gm-c1080"

/note="Vector: pBluescript II SK+, Site 1: EcoRI, Site 2: XhoI; The RNA was isolated from roots of a 8 day old 'Brags' supermodulating mutant NTS382 seedlings that were infected with Bradyrhizobium japonicum, strain USDA 110, 72 hours prior to harvest. Dr. Gary Stacey generously donated the tissue. The roots were flash-frozen in liquid nitrogen. Stragene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stragene's first-strand synthesis primer was used. An 'anchor' nucleotide (V-A, C, or G) was added to the 3' end of the primer (GAGAGAGAGAGAGAGACTAGCTCGAG(T)18V) to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The cDNA was then precipitated and redissolved in sterile, RNase-, DNase-free water. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI from Promega (400U/ul), all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using Sephacryl S-500 High Resolution (Pharmacia Biotech) in a 2-mm diameter column and a bed volume of approximately 1ml. The column eluent was precipitated, redissolved, and ligated into Stragene's pBluescript II XR Predigested vector (pBluescript II SK(+)) vector that has been digested with EcoRI and XhoI, and phosphorylated by Stragene). This library was constructed in the laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at Northern Arizona University."

ORIGIN

Query Match 7.4%; Score 33.8; DB 13; Length 559;

Best Local Similarity 55.6%; Pred. No. 1.2e+02;

Matches 65; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 76 TGGGGCGGGCGGGTATCATTAACGGCGGCAATAGTCCGGCCGGACTACAGTTG 135

Db 233 TGGGGTCAGAGGACCCCATACAGTCCGGCAACATAGTCCGGACATGTGTG 174

Qy 136 AGCATTTATAGTACGTTCCGCTTAACGCTGCTGCTGCAAGAGTCCCGT 192

Db 173 AAGATGCCCAAGTACGTTGATGATGAGGACGCTTTGAAGCCAAAGATGCTGCCGT 117

RESULT 13

LOCUS A0637195/C

DEFINITION RPCI-11-481M9.TV RPCI-11 Homo sapiens genomic clone RPCI-11-481M9, genomic survey sequence.

ACCESSION A0637195

VERSION A0637195.1 GI:5099830

KEYWORDS GSS.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 695)

Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P., and Venter, J.C.

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

Unpublished (1997)

Other GSSs: RPCI-11-481M9.TV

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@reggen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search.html.

Seq primer: T7

Class: BAC ends.

Location/Qualifiers

1. 695

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="GDB:7684616"

/db_xref="taxon:9606"

/clone="RPCI-11-481M9"

/sex="Male"

/cell_type="lymphocytes"

/clone_id="RPCI-11"

/note="Vector: pBAC3.6; Site 1: EcoRI, Site 2: EcoRI; RPCI11 Human Male BAC Library"

ORIGIN

Query Match 7.4%; Score 33.8; DB 28; Length 695;

Best Local Similarity 49.2%; Pred. No. 1.4e+02;

Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

Qy 275 TGACTCAGATGTTTCAGAAATTAATGACCACTGACGATGGAAGCTTAAACCTCG 334

Db 291 TAACATATATCTTATACACACAAAAGAGATTAATGATCAAGTCTAGACACGT 232

Qy 335 ATATTACTGTCGGCAATACGCGGATATACGCGGCTGTTAATTATGATCAGCTG 394

Db 231 ATTTACAGTTTATACCAAGTACATGATGACATGACAGCTGTACATCATACTACAG 172

Qy 395 TTAACCGTGTGTTTACCATGAATGACATGCAACAAAGCCAGCGCTAACCACTATT 454

Db 171 CCATACATGTGACACACATTAAGATCAATTAAGATGACACAAAGTATACATTCAT 112

OY 455 A 455
Db 111 A 111

RESULT 14

CG178280 941 bp DNA linear GSS 21-AUG-2003
LOCUS CG178280
DEFINITION PUJDF797D_ZM_0.6_1.0_KB Zea mays genomic clone ZM82A0641N14,
genomic survey sequence.
ACCESSION CG178280
VERSION CG178280.1 GI:34069341
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE
AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
1 (bases 1 to 941)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uterback, T.,
Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.

TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUJDF797B
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@cigr.org
Seq primer: 7P
Class: sheared ends.

FEATURES

source
1..941
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM82A0641N14"
/clone_11b="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TORO; Site_1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"

ORIGIN

Query Match 7.4%; Score 33.8; DB 29; Length 941;
Best Local Similarity 51.0%; Pred. No. 1.6e+02;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
OY 17 TGGAGCATTCGCGCATCTGTTCTTGGCAGTCTCTGGCTGGCGTTCACAAAT 76
Db 236 TTGAGCTCTACATCAGTGTCTATATGCAAGTCTCTGCAACCGGCTTCCGCGG 295
OY 77 GGGCGCGCGGATATCATTAACGGCGGCAATAGTTCGCGCGGANTCAAGTTGA 136
Db 296 GGAATACCTCCGTCGTCACACGACCGCGGCTTGAAGTTCCGCGACTGTAACCGCA 355
OY 137 GCATTTATCAGTACGTTCCGCTTAAGCGTGGCTTGC 173
Db 356 CTCTCTTCGCAACCGCTCAACGCGACGACTCTGGC 392

RESULT 15

CNS02G0P/c 1098 bp DNA linear GSS 01-SEP-2000
LOCUS CNS02G0P/c
DEFINITION Tetradodon nigroviridis genome survey sequence PUC-ORI end of clone
138K23 of library G from Tetradodon nigroviridis, genomic survey
sequence.

ACCESSION AL196666
VERSION AL196666.1 GI:7834816
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradodon nigroviridis
ORGANISM Tetradodon nigroviridis

REFERENCE

AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brothier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetradodon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645

TITLE

Roest Crolius, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C.,
Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradodon nigroviridis

Genome Res. 10 (7), 939-949 (2000)

JOURNAL

COMMENT

Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefgenoscope.cns.fr)
- Web : www.genoscope.cns.fr
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradodon.

FEATURES

source
1..1098
/organism="Tetradodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="138K23"
/clone_11b="G"
/note="Genoscope sequence ID : CNS0138AF12SP1-end ;
PUC-ORI"

ORIGIN

Query Match 7.4%; Score 33.8; DB 29; Length 1098;
Best Local Similarity 55.8%; Pred. No. 1.7e+02;
Matches 53; Conservative 5; Mismatches 37; Indels 0; Gaps 0;
OY 19 GCAGCATTCGCGCAATGCTGTTCTTGGCAGTCTCTGGCTGGCGTTCACAAATG 78
Db 773 GAACTCCGCTGAACAGGGGGAATTTGGCGTAAAGCTGCTGAATGTTCAACGAT 714
OY 79 GCGCGCGCGGTAATCATTAACGGCGCGCGCAATG 113
Db 713 AAAGCGCGCGTTTAACTTMAAGCCGACGCGCAATG 679

Search completed: March 16, 2004, 04:29:02
Job time : 2232.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456

Sequence: 1 atgaactttaaagtcgc.....ccacgctaacagctactaa 456

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Genembi:*

- 1: gb_ba:*
- 2: gb_bcg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_vl:*
- 14: gb_vl:*
- 15: em_da:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_ov:*
- 22: em_ov:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hcg_hum:*
- 31: em_hcg_inv:*
- 32: em_hcg_other:*
- 33: em_hcg_mus:*
- 34: em_hcg_pln:*
- 35: em_hcg_rtd:*
- 36: em_hcg_mam:*
- 37: em_hcg_vtc:*
- 38: em_sy:*
- 39: em_hcgo_hum:*
- 40: em_hcgo_mus:*
- 41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	404.8	88.8	456	144909	144909 Sequence 58
2	404.8	88.8	2067	1 SEU43280	U43280 Salmonella
3	396.8	87.0	5103	1 STAJ2301	AJ002301 Salmonella
4	396.8	87.0	22411	1 AE008749	AE008749 Salmonella
5	395.2	86.7	254050	1 AL627269	AL627269 Salmonella
6	395.2	86.7	301983	1 AE016840	AE016840 Salmonella
7	389.4	85.4	1048	1 STAFBFA	AU000514 Salmonella
8	287.2	63.0	361	6 144908	144908 Sequence 56
9	281.6	61.8	2889	1 CSP515700	AJ515700 Citrobact
10	229.2	50.3	2920	1 CFR515701	AJ515701 Citrobact
11	229	50.2	230	1 SEU53207	U53207 Salmonella
12	227.2	49.8	4680	1 ECCSGABDC	X90754 E.coli csgG
13	227.2	49.8	10346	1 AE000205	AE000205 Escherich
14	227.2	49.8	15047	1 D90741	D90741 Escherichia
15	225.6	49.5	456	6 AX814811	AX814811 Sequence
16	225.6	49.5	648	1 ECOCGGA	L04979 Escherichia
17	220.6	48.4	306358	1 AE016759	AE016759 Escherich
18	217.4	47.7	10190	1 AE005315	AE005315 Escherich
19	217.4	47.7	10190	1 AP002554	AP002554 Escherich
20	217.4	47.7	327773	1 ESAS15702	AJ515702 Enterobac
21	188.8	41.4	2883	1 AE015131	AE015131 Shigella
22	163.2	35.8	10370	1 AE016981	AE016981 Shigella
23	163.2	35.8	292504	1 AE016981	AE016981 Shigella
24	140.2	30.7	437	1 AP237726	AP237726 Shigella
25	64.6	14.2	1212	1 EC0131756	AJ131756 Escherich
26	64.4	14.1	19201	1 D90742	D90742 Escherichia
27	44.8	9.8	72	6 AX814798	AX814798 Sequence
28	44.8	9.8	2000	6 AX655393	AX655393 Sequence
29	39.4	8.6	956	8 AK058943	AK058943 Oryza sat
30	39.4	8.6	2720	8 AK102629	AK102629 Oryza sat
31	39.4	8.6	152448	8 AP004339	AP004339 Oryza sat
32	38.8	8.5	13264	1 AE001012	AE001012 Archaeogl
33	38.2	8.4	125020	9 AF429315	AF429315 Homo sapi
34	36.2	7.9	1425	1 AF343445	AF343445 Lactobaci
35	36.2	7.9	302156	3 AC116977	AC116977 Dictyoste
36	36.2	7.9	168267	3 AC116977	AC116977 Pan trogl
37	35.8	7.9	172600	9 AL359644	AL359644 Human DNA
38	35.8	7.9	193168	2 AC011969	AC011969 Homo sapi
39	35.6	7.8	301214	1 AE016786	AE016786 Pseudomon
40	35	7.7	117541	2 AC104285	AC104285 Oryza sat
41	34.4	7.5	560	8 AY247273	AY247273 Medicago
42	34.4	7.5	1553	1 HAAR165	MS9123 Haemaphysob
43	34.4	7.5	2751	6 AX416572	AX416572 Sequence
44	34.4	7.5	151828	2 AC146008	AC146008 Pan trogl
45	34.4	7.5	160449	10 AC130549	AC130549 Mus muscu

ALIGNMENTS

RESULT 1

LOCUS 144909

DEFINITION Sequence 58 from patent US 5635617.

ACCESSION 144909

VERSION 144909.1 GI:2469622

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 456)

AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.,

TITLE Methods and compositions comprising the agfa gene for detection of

Salmonella

JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

FEATURES

Source 1: 456
Location/Qualifiers
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match

Best Local Similarity 88.8%; Score 404.8; DB 6; Length 456;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTAAAGGCGCAGCATTTGCGAGCAATGATGTTCTGGCAGTCTGGCT 60
DB 1 ATGAACTTTAAAGGCGCAGCATTTGCGAGCAATGATGTTCTGGCAGTCTGGCT 60
QY 61 GCGCTCTTCCACATGCGCGCGCGGCGGATTCATACGCGCGCAATAGTTCGCG 120
DB 61 GCGCTCTTCCACATGCGCGCGCGGCGGATTCATACGCGCGCAATAGTTCGCG 120
QY 121 CCGGACTCAACGCTGAGCATTTTACATGACGTTCCGCTAACGCTGCGCTTGCCTGCA 180
DB 121 CCGGACTCAACGCTGAGCATTTTACATGACGTTCCGCTAACGCTGCGCTTGCCTGCA 180
QY 181 AGCGATGCCGTAATCTGAAACGACATTCACGAGCGGTTATGCTAAACGCGCGAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACATTCACGAGCGGTTATGCTAAACGCGCGAT 240
QY 241 GTAGGCGCAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGCAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCGACCATGACGACGAGTGAACGCTAAACCTCCGATTTCTGCGCAATATGATCAG 360
DB 301 GCGACCATGACGACGAGTGAACGCTAAACCTCCGATTTCTGCGCAATATGATCAG 360
QY 361 CTGTTACCGCGTGTGTTACCATGAATGCGACATGCGATGATGCGTGGCGGAT 420
DB 361 AATAACCGCGCGTGTGTTACCATGAATGCGACATGCGATGATGCGTGGCGGAT 420
QY 421 GGTTTGGCAACACGCGACGCGCTAACGATTTAA 456
DB 421 GGTTTGGCAACACGCGACGCGCTAACGATTTAA 456

```

RESULT 2
LOCUS SEU43280 2067 bp DNA linear BCT 14-FEB-1996
DEFINITION *Salmonella enteritidis* agfBAC operon: fimbria-like protein
precursor (agfB), thin aggregative fimbriae precursor (agfA), and
AgfC (agfC) genes, complete cds.
VERSION U43280.1 GI:1184712

ACCESSION U43280
KEYWORDS
SOURCE
ORGANISM *Salmonella enteritidis*
Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; *Salmonella*.

REFERENCE 1 (bases 1 to 2067)
AUTHORS Doran, J.L., Collinson, S.K., Burian, J., Sarlos, G., Todd, B.C.,
Munro, C.K., Kay, C.M., Baner, P.A., Peterlin, P.I., and Kay, W.W.

TITLE DNA-based diagnostic tests for *Salmonella* species targeting agfA,
the structural gene for thin, aggregative fimbriae
JOURNAL J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
MEDLINE 94013373
PUBMED 8104955

REFERENCE 2 (bases 1 to 2067)
AUTHORS Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and
Kay, W.W.

TITLE *Salmonella enteritidis* agfBAC operon encoding thin, aggregative
fimbriae
JOURNAL J. Bacteriol. 178 (3), 662-667 (1996)
MEDLINE 96146512
PUBMED 8550497

REFERENCE 3 (bases 1 to 2067)

FEATURES

Source

Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
Direct Submission
Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
Microbiology, BC V8W 3P6, University of Victoria, P.O. Box 3055 Petch Bldg.,
Victoria, BC V8W 3P6, Canada
Location/Qualifiers
1: 2067

-35_signal
-10_signal
gene
CDS
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrc"
571..576
598..603
696..1151
/gene="agfB"
696..1151
/gene="agfB"
/note="fimbria-like protein precursor"
/codon_start=1
/product="agfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKNLKLPMLTILGAPGATATNTVLDLSEVNFVNLKSPFN
SISQSAVNSAIIQKSGNKNVITGTQKTAIVVQKSHAIRVQR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
/product="agfB"
/note="putative signal sequence of 21 amino acids;
fimbria-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbria subunit precursor; major
fimbrial subunit of thin aggregative fimbriae precursor"
/transl_table=1
/product="AgfA fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLKVAPFAIVSGSALAGVVPQWGGGNNNGSGSPDST
LSIYQSGANALALQSDARSEPTTOSGCGADVGADNSTIELTONGFRNAT
IDQWNAKSDITVGGYGNNAALVNTASDSVVRQVGGNNAATNQQ"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/gene="agfA"
/product="AgfA fimbria"
/note="thin aggregative fimbria subunit; major fimbrial
subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=1
/product="agfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MHTLLLAASNDITFTTQGGDIYVIVPQTLNRCVQVOYL
SYRDVGCGSHRQKQTLSPANQPIELSRVLSVSSDSVKLIIVYDSQSLHLSQO
WPSAQ"

ORIGIN

Query Match
Best Local Similarity 88.8%; Score 404.8; DB 1; Length 2067;
Matches 93.0%; Pred. No. 1e-105;


```

gene
  3344..3376
  /gene="csGBA"
  /protein_id="CA05319.1"
  /db_xref="GI:2739241"
-10_signal
  3344..3349
  /gene="csGBA"
  3371..3376
  /gene="csGBA"
  3469..3924
  /gene="csGBA"
  3469..3924
  /gene="csGBA"
  3469..3924
  /gene="csGBA"
  /function="nucleator"
  /codon_start=1
  /evidence=experimental
  /transl_table=1
  /product="CsgB protein"
  /protein_id="CA05316.1"
  /db_xref="GI:2739238"
  /db_xref="GOA:P55226"
  /db_xref="SWISS-PROT:P55226"
  /translation="MKKKLFLMMLTGIATPTDARSEYFNAVELSKSFN
  QAAITGVGTDNARVROGSKLVTISGEGNNRKAVDGANNFAYIEOTGNADA
  S1SGASGNSAAAIQKSGNKANITQYGTQKTAIVVQKSHMAIRVQR"
  3469..3531
  /gene="csGB"
  3532..3921
  /gene="csGB"
  /product="CsgB protein"
  /function="nucleator"
  3966..4421
  /gene="csGA"
  3966..4421
  /gene="csGA"
  /function="fiber subunit"
  /codon_start=1
  /evidence=experimental
  /transl_table=1
  /product="CsgA protein"
  /protein_id="CA05317.1"
  /db_xref="GI:2739239"
  /db_xref="GOA:P55225"
  /db_xref="SWISS-PROT:P55225"
  /translation="MKLKVAAPAIVSGSALAGVFPOMGGGNGHGGSSPST
  LSIYQGSAAALAIQSGARSETTISGNGADVGAGADNSTIETLNGRRNAT
  IDWNAKNSDITVGTQGNNAALVNOTASDSVWVRQVFGNATANQY"
  3966..4025
  /gene="csGA"
  4026..4418
  /gene="csGA"
  /product="CsgA protein"
  /function="fiber subunit"
  4483..4809
  /gene="csGC"
  4483..4809
  /gene="csGC"
  /codon_start=1
  /transl_table=1
  /product="CsgC protein"
  /protein_id="CA05318.1"
  /db_xref="GI:2739240"
  /db_xref="SWISS-PROT:P55227"
  /translation="MHTLLIALASKQITFTTQGGDIYTVIPQVTLNPEQVOYIL
  SVRDVGQGGSHQKQTLPLPAPQPIELSRSLVINSRDSVKIIVTSDQSILHSQ
  WPSAQ"
  4483..4533
  /gene="csGC"
  4534..4806
  /gene="csGC"
  /product="CsgC"
  4940..5103
  /gene="ORF103"
  4940..5103
  /gene="ORF103"
  /codon_start=1

```

```

ORIGIN
  /transl_table=1
  /protein_id="CA05319.1"
  /db_xref="GI:2739241"
  /db_xref="SWISS-PROT:O54295"
  /translation="MSVKKKIPALIGLCIAFPIHSAVQGVQGVHFRGAIPEPL
  CDSIYHANI"

```

```

Query Match      87.0%  Score 396.8, DB 1, Length 5103;
Best Local Similarity 91.9%, Pred. No. 2,3e-103;
Matches 419; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGCAGCATTCGACGATCGTGGTTCTGCGACGCTCTGGCT 60
Db 3966 ATGAACCTTTTAAAGTGCAGCATTCGACGATCGTGGTTCTGCGACGCTCTGGCT 60
QY 61 GCGCTCGTTTCAACATGGGCGGCGGCGGTATTCATTAACGCGCGGCAATAGTTCCGGC 120
Db 4026 GCGCTCGTTTCAACATGGGCGGCGGCGGTATTCATTAACGCGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGACCTTTATAGTACGTTCCGCTAACGCTGGCTTCTTGCAA 180
Db 4086 CCGGATTCACGTTGACCTTTATAGTACGTTCCGCTAACGCTGGCTTCTTGCAA 180
QY 181 AGCGATGCCGTTAACTGAAACGACCATTAACCGAGCGGTTATGTAACGCGCGCAT 4145
Db 4146 AGCGATGCCGTTAACTGAAACGACCATTAACCGAGCGGTTATGTAACGCGCGCAT 4145
QY 241 GTAGCCAGGCGGTGGGAAATAGTACTATTAAGTAACTGAAATGTTTCAAAATAT 300
Db 4206 GTAGCCAGGCGGTGGGAAATAGTACTATTAAGTAACTGAAATGTTTCAAAATAT 300
QY 301 GCCACCATTCGACGATGGAACGCTAAAACTCCGATATACGTGCGCAATATGATCAG 360
Db 4266 GCCACCATTCGACGATGGAACGCTAAAACTCCGATATACGTGCGCAATATGATCAG 360
QY 4266 GCCACCATTCGACGATGGAACGCTAAAACTCCGATATACGTGCGCAATATGATCAG 360
Db 4326 AATTAACCGCGCGGTGTTATGACACCGCATCTGATTCACGCTATGTCGCTCAGGTT 420
QY 4326 AATTAACCGCGCGGTGTTATGACACCGCATCTGATTCACGCTATGTCGCTCAGGTT 420
Db 421 GGTTCGGCAACACCCGCGCTAACGATTTAA 456
QY 4386 GCTTTGGCAACACCCGCGCTAACGATTTAA 4421
Db 4386 GCTTTGGCAACACCCGCGCTAACGATTTAA 4421

RESULT 4
AE008749 22411 bp DNA linear BCT 23-APR-2003
LOCUS AE008749
DEFINITION Salmonella typhimurium LT2, section 53 of 220 of the complete
genome.
ACCESSION AE008749 AE006468
VERSION AE008749.1 GI:16419641
KEYWORDS
SOURCE
ORGANISM
Salmonella typhimurium LT2
Salmonella typhimurium LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 (bases 1 to 22411)
McLellan, P., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Hou, S., Layman, D., Leonard, S., Portnoy, L.K., Ali, J., Dante, M., Du, F.,
Grewal, N., Mulvaney, E., Leonard, S., Nguyen, C., Scott, R., Holmes, A.,
Stonington, T., Nhan, M., Waterston, R. and Wilson, R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE 2 (bases 1 to 22411)
AUTHORS
TITLE
The Salmonella typhimurium Genome Sequencing Project
CONSTRM Direct Submission
JOURNAL Submitted (29-MAR-2001) Genome Sequencing Center, Department of

```

Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108 USA
COMMENT Supported by NIH grant 5U 01 AI43283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-VIDES and
RegunoldB;
http://kinich.cifn.unam.mx:8850/db/regunoldb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m3 subclone.

FEATURES
SOURCE location/Qualifiers

1..22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
434..1308
/gene="phoH"
/note="synonym: STM1126"
434..439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"
434..1308
/note="similar to E. coli phoB-dependent, ATP-binding pho
regulon component; may be helicase; induced by P
starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=1
/product="phoB-dependent ATP-binding pho regulon
component"
/protein_id="AA120057.1"
/db_xref="GI:16419642"
/translation="MCANRRVQENAGDVYVQLKLSNGRQAVIKAREAKVLRDS
RSHKQREESVTSLVQGGVEALIGMARDSDRTSVKARNEQAHYLNAIDSKOIFAT
GEACGKTWISAAKABALIKHDERIIVTRPVQADQEDLGLPDIAEKAPYFRFV
YDVLKRLGASFMQYCLRPEIGKEIAPFAVGRGTENAVVILDEAONTVAOMKRF
LTRIGENVTVIVNDITQCCLPRGVRSGSLDALRFEDEWGIIVHNKDCVNSALC
ORTIHAYS"
complement(1414..2303)
/gene="STM1127"
complement(1414..2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
identity in aa 15 - 264"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="AA120058.1"
/db_xref="GI:16419643"
/translation="MEQPPRLKPKGLIDLTIGMOKSITRASQRIAGYIAPFQVTO
SSINDISRETQAGEATVIRFRTLVGYGQDFKNDLIELATTSDDSSPLDVEE
SDDHAIGLKIONTSIVLSETNLIDMOQLGVADLRHCHSYVIFGVSSGTTAD
MKHLMRIGLRGDAVSNHFMWQATLLKAGDVAAGVSHSGTSPETVHSIRLAQACA

TTVAITNIGSPICEADFCILINRGOMLOGDSIGTKAQLFVFDLLYTLVQSSPE
OARSKXIRTNALDMTK"
complement(2298..2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"
complement(2580..4088)
/gene="STM1128"
complement(2580..4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=1
/product="putative sodium/glucose cotransporter"
/protein_id="AA120059.1"
/db_xref="GI:16419644"
/translation="MITHSRGIVVYLVLEFYLAMLVGVYFSRROKTDADYFGGR
VPGMAGVSPATLTLSITPMSIPAKFTSDWPFILGOYLAIALIPLVFPYIPERK
LKTSAYEYLRARDVRCRLPASRPMLEPHGRIITFLVLAIRPIAIDPILVL
LISVLCITVYMWGIGEVITVDVIOGLLSALILITVLCVKQGGIDELFTVQA
DKFPPATQFMSWTSETPVLMIGFLPANTIQOFPASQDVQRYVTDISETKTLT
NAKLAVIIPVFPALGALFVYQOHPQLPAGFTGILPLFVTEMPVGIAGLIA
AIPAAQSSISSINSISCFNSDIYORLSHKRTPENRMKILAVILVAGLISAA
VLMVADSESHMDAFNSLIGMGPMGLPMLGIFFRANAGSAVIGITISVITVGA
RYATDLNPFYGVIGSLSVITSGYIFRPLPAPPLTLDEKPEKRTL"
complement(4083..4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"
complement(4413..5099)
/gene="STM1129"
complement(4413..5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AA120060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSHENGGLIVSCQPVGSPMDKPRIAAMAQAA
AGAAVAREIGENLRTPHLSVPIIGIRLDLGSRTIPRYQVDALAQADII
APGAVFRSRVDLSILTRIRHGLTAMADCSYNEISCHQKIEFGITLSYGTQ
ITPVEDPLAVTQLSHAGRCVIAEGRYNTPLAANALRHGAIVAVTSALIRISHIQ
WFSHAVKR"
complement(5094..5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"
5599..6759
/gene="STM1130"
5599..6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AA120061.1"
/db_xref="GI:16419646"
/translation="MGQMKNPKKMTLALCSVAITTSQVATLLPDIPPLKNG
AIDNNGVITYGLGTAGSYTKIDKQKHMERIKSPGAGREQSVFLADELTVFG
GVGRKNSSEPLQVSYDVYKYSPVKNMVKQDITSPVGLTGTGKRELSVTLTGAVN
EHI.FDKYFIDIAAADADESEKNKVIYVFKPAKADYEFNKIVIPYNAKENTWAGEL
PGAGTQSSSVMGNNFLMLINGELKPLRFVDVLYRAMMDKLTMLNQSOLPSPGEO
QOGLGAFSGYSHGVLLVGGANFPKAPKQNTYNGKXSHSGINKKRDVYGLINCH
MOGWMKQKQBLIGVSVSYGVDFVLIGENAKKAPSVSVSPMRDNLILIK"
6793..7497
/gene="STM1131"
6793..6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

CDS

6805..7497
 /gene="STM1131"
 /note="similar to E. coli orf, hypothetical protein
 (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%

identity in aa 4 - 239"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AA020062.1"
 /db_xref="GI:16419647"
 /translation="WKINKYLQVSEFIAFSSYLQATLDYHEVADRTINCRRI
 IELKPNGIGFVDSVKSVDGEODKHLSDIVANAIEGVSNKATDNLQPGRI
 DESGPDISIKPYIRGOYNDSGVMAAGRYDPAKRNYSDBKINRFDPTYGIF
 DELKLEYNFTMDSDDIKFDNKKTEHNVALAMKLNSFTPYVEGVAVRNNTDER
 QTRRYVLEQYHF"
 7768..9060
 /gene="STM1132"
 /note="STM1132"
 7768..7773
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
 7780..9060
 /gene="STM1132"
 /note="similar to E. coli putative transport protein
 (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%

identity in aa 25 - 418"
 /codon_start=1
 /transl_table=11
 /product="putative sugar transport protein"

Query Match 87.0%; Score 396.8; DB 1; Length 22411;
 Best local similarity 91.9%; Pred. No. 2.6e-103;
 Matches 419; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCAGCATTCGCGCAATCGTATGTTCTTGCGAGTCTCTGCT 60
 DB 17769 ATGAACTTTTAAAGTGCAGCATTCGCGCAATCGTATGTTCTTGCGAGTCTCTGCT 17828
 QY 61 GCGCGCTTCCACATGATGGCGCGCGCGGCGGTAATCATAGCGGCGGCAATAGTTCCGGC 120
 DB 17829 GCGCTCTTCCACATGATGGCGCGCGCGGTAATCATAGCGGCGGCAATAGTTCCGGC 17888
 QY 121 CCGGATCGAAGTTGAGCATTTATCGATAGGTTCCGCTAACGCTGCGCTTCTGCTGCA 180
 DB 17889 CCGGATTCGAGTTGAGCATTTATCGATAGGTTCCGCTAACGCTGCGCTTCTGCTGCA 17948
 QY 181 AGCGATGCCCGTAAATCTGAAACGACCATTAACCGAGCGGTTATGTTAAGCGCGCGCAT 240
 DB 17949 AGCGATGCCCGTAAATCTGAAACGACCATTAACCGAGCGGTTATGTTAAGCGCGCGCAT 18008
 QY 241 GTAGCGCAGGTTGCGGATTAATGTAATCTTGAATCTGAGAAATGTTTCAAAATAT 300
 DB 18009 GTAGCGCAGGTTGCGGATTAATGTAATCTTGAATCTGAGAAATGTTTCAAAATAT 18068
 QY 301 GCCACCATGACCAAGTGAACGCTAAAACTCCGATTTACTCGCGCAATATATGATCAG 360
 DB 18069 GCCACCATGACCAAGTGAACGCTAAAACTCCGATTTACTCGCGCAATATATGATCAG 360
 QY 361 CTGGTATCCGCTGTTTATCCCATGGAATGCGACATGCAAGCGTAAATGTCGTCAGGTT 18128
 DB 18129 AATAACCGCGCGGCTTATATGACCGCATTCGATTCAGCGTAAATGTCGTCAGGTT 420
 QY 421 GGTTTGGCAACAACGCGCAGCGCTAACCGATTTAA 456
 DB 18189 GGTTTGGCAACAACGCGCAGCGCTAACCGATTTAA 18224

RESULT 5
 AL627269 254050 bp DNA linear BCT 04-JUL-2003
 LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
 DEFINITION Complete chromosome, segment 5/20.
 ACCESSION AL627269 AL513382
 VERSION AL627269.1 GI:16502231

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCES

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

misc_feature

misc_feature

Salmonella enterica subsp. enterica serovar Typhi
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 1 (bases 1 to 254050)
 Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D.,
 Sebitiha M., Baker S., Basham D., Brooks P., Chillingworth T.,
 Farrar J., Feltwell T., Hamlin N., Haque R.M., Dowd L., White N.,
 Jagers K., Krogh A., Larsen T.S., Leach S., Hien T.T., Holroyd S.,
 Parry C., Quail M., Rutherford K., Simmonds M., Skelton J.,
 Stevens K., Whitehead S. and Barrett B.G.
 Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18
 Nature 413 (6858), 848-852 (2001)
 21534947
 11677/608
 2 (bases 1 to 254050)
 Parthill J.
 Direct Submission
 Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
 sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, UK
 E-mail: parthill@sanger.ac.uk
 Notes:
 Details of S. Typhi sequencing at the Sanger Centre are available
 on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
 Location/Qualifiers
 1..254050
 /organism="Salmonella enterica subsp. enterica serovar
 Typhi"
 /mol_type="genomic DNA"
 /strain="CT18"
 /db_xref="taxon:90370"
 181..2793
 /gene="STY1078"
 /note="synonym: pepN"
 181..2793
 /gene="STY1078"
 /note="Orthologue of E. coli pepN (AMPN ECOLI); Fasta hit
 to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="aminopeptidase N"
 /protein_id="CAD08183.1"
 /db_xref="GI:16502232"
 /db_xref="GOA:Q82770"
 /db_xref="SPTRMBL:Q82770"
 /translation="MTQDPQAKYRHYRDPYQITDIELTDLDAEKTVVTAISQAVR
 HGARDPRLDGEDLTVLSIHVADAEMTVKKEBGLITISDIPERFTRITVNEISPA
 NTALEGLYQGDALCTQCEAGEGRHTTWLDRPVLARTTKIDKSKYFELLVNSPA
 RVAQGLENRHNVQWQDPEPKCYLFAVADPVLADPTTSGRVALVELYVDSG
 NLDGAPAMTSLAKSMKMDETRPGLEYDIDITVAVDPEFMGAMENGLVIFSKYV
 LARDTADTDYLDIERVIGHEYPFHVGNRKTVCIDWFOLSIKGLTVFRDQESSDL
 GSAVRVRSINVRMTRGLOPRAEDASPAHPIRDKVIEKNNTYTLVVEKGEVIMRW
 TLLGEENFOKGMOLYERHDSAAATCDPEVAMEDASVDSHFRMYSSGDPITVY
 VNVTQABQTFEDVNVYRQPPALCEFSAPVKLEVKKSQDQELFLMRARNDPSRD
 AASGLATYIKLVNARHQOGPLSLPVHADAFAVAILDEKIDPALAAELTLPSANE
 IAEFLFEVIDPIAIAQVREALTRTLAELADEFLAIYVANHDEYKVDHGDIGRTLDN
 ACIRPLTFGEETLANTLVSKYRDANNMTDIALSAVAVAQPCBDTLMQAEYDDKM
 ODGLVMDKWFILQSTSPAEVNTLVRSILKRSFSMSNPRIISLIGAFSGSPAAHF
 AODGSGYQVLVMTLDLNSRNPVASRLIEPLIRLKRYDDKQEKRAALEQKLEN
 LSGDYEKITKTAIA"
 532..1332
 /gene="STY1078"
 /note="Pfam match to entry PF01433 peptidase_M1, Peptidase
 family M1, score 245.10, E-value 1.9e-84"
 1060..1089
 /gene="STY1078"

/note="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
/gene="STY1079"
3001..4011
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit
to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q82789"
/db_xref="Swiss-Prot:Q82789"
/translation="MYPFVKALFQDLPRAHEFTFOQLRITGTPLEALVROKVP
KPYCMGLTFKPLGLAAGLDKCEIDALGAMGSLGISTVTPRPOGNDKRLPR
LVDAEGLINRGFNGLGVNLEVENKKAHFGIIGINIGKNDKOTPVENGKDYLCME
KVYAVGYIAINISSPMTPLGLTLYDGLDLDLITAIKNKNDLOVHHKYPVAVKI
APDCERELIOVADSLRNITDVIATNTTIDRSLVQMKRCQOTGGLSGRPLDKST
EIRRLSELKQLPIIGVGIDSVIAREKIAAGATLVQISGPIFGPPLKEIYV
HI"
3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODHase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q82758"
/translation="MMRIKPDNNRWYVDEEDHDMMLDLANGMLFRSFRSKMLTPDA
FCTGFCVDAAALYFSEFKCRDELTKEORALVINALVALIRYLKPOMKSMHFAVA
GEMWTGCTGAASVMSDPTAEQVNLVYVERENALCLLAPGVIVIGRTMGLDAIK
IMNDRLKPOVHCSPFLEQAV"
complement(4716..5825)
/gene="STY1081"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82787"
/db_xref="SPTREMBL:Q82757"
/translation="MATLSLFIHPIVPSKMGIGLTHALADISGLAFRIMVETSDGT
FITAPQPMWRFTPSLHDGLHLTPADSSALVRFDPDPTFMVGNHFAARA
PTAINVLSEGFSDVOLCWVGPOLTRVYKSHNAVPLGDFPDGYLYLTNNSLSRLDLO
RCAGVQWQGFPRMLVYSGVAWEEDSMKYLRTGDIYDVYKPCSRCITPTNVEPKCO
KH8SGPLATLQARTNODNGVDVDFGNLLARNSGVRVDEVEITLTAAPAKAYGATT
LDGVSVEKHPDGSVITDQSGITFCGNNOVLLEQLNNGIRLIPYSCRAIGCGCRIR
LDSGEVSPKRSKMGWDGTTILSCSVCKTALRLEN"
complement(4737..4955)
/gene="STY1081"
misc_feature

/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical
protein h10116 H10116 SW:YCBY_HAEIN (P44524; P43945) Faeta
scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82786"
/db_xref="SPTREMBL:Q82786"
/translation="MNSLPASTARGLEELIKLEKAGVCCOVVQGVHRCQTRLI
YOSLMSRLSRILPEWCKVSDLDLYFGVQXINTEIFNPGATVAVHPSGLNDTI
RNSQYGMKVKDAI VDAFTKRLPRPVNDESDPLRIWMLNKETASIALDSDGLH
LRGYRDRTGAPIKETLAAAIIVMSGQPTPLIDPMCGSGTLLIEAAMVADRAPGL
HRGHGFGSGMAOHDEAITWQVKAAQOTRAKGLAEYSHFSGSDARVERASNNR
RAGIGELITEFVNDVQLSNPRLKPGYGVISNPVGERLSDSPALIALSLIGRTMK
NORGWMLSPFASPDLSLOLRADQOFKAKNGPLDVCYKNTIATTTADSKRATYA
EDIANRLRKLKLEKARQEGIECYRLVADLPETVAVADRQDMVAVIYAPKTV
DAQKARQLFDIIAATLSVIGIPNNKVLITREOKKNQYOKNSEGEFLVSEYNA
RLWNLITDYLDITGLFDHRIARMLGMSGXKGFNLFSYTSASVAGSGAGANTTT
VDMSTRYLEWAERNLRLNGSGRAHRLIADCGMLAEANEQPLFIPIPTPFNSNR
MEESPVDYORHDIWMLDKRLRLKKGITIMSNKRGFRMDLEGLAEGLTAQETTKT
LSPDFARNRQIHNCMLIRAA"
5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJUX_ECOLI (554 aa), 34% identity in 524 aa
overlap
Query Match 86.7%; Score 395.2; DB 1; Length 254050;
Best Local Similarity 91.7%; Pred. No. 9.6e-103;
Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGAGCATTCGGAGCAATGCTAGTTCTGACAGTCTGGCT 60
DB 88906 ATGAACCTTTAAAGTGGAGCATTCGGAGCAATGCTAGTTCTGACAGTCTGGCT 88965
QY 61 GCGGTCGTTCCACAAATGGGCGGGCGGTATCATATACGGCGGCAATAGTTCCGGC 120
DB 88966 GCGGTCGTTCCACAAATGGGCGGGCGGTATCATATACGGCGGCAATAGTTCCGGC 89025
QY 121 CCGGACTCAACGTTGAGCATTTATCAATACGTTCCGCTAACGCTGCGCTTCTGCAA 180
DB 89026 CCGGACTCAACGTTGAGCATTTATCAATACGTTCCGCTAACGCTGCGCTTCTGCAA 89085
QY 181 AGGATGCGCGTAAATCGTGAAGGACATTAACCAAGGCGGTTATGTTAAGGCGCGCAT 240
DB 89086 AGGATGCGCGTAAATCGTGAAGGACATTAACCAAGGCGGTTATGTTAAGGCGCGCAT 89145
QY 241 GTAGGCGGAGGTGCGATTAATAGTACTATTGAACCTGACAGATGTTTTCAGAAATTAAT 300
DB 89146 GTAGGCGGAGGTGCGATTAATAGTACTATTGAACCTGACAGATGTTTTCAGAAATTAAT 89205


```
gene
      complement(5597..6295)
      /gene="figh"
      /locus_tag="t1740"
      complement(5597..6295)
      /gene="figh"
      /locus_tag="t1740"
      /note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
      /codon_start=1
      /transl_table=11
      /product="flagellar L-ring protein precursor"
      /protein_id="AA069364.1"
      /db_xref="GI:29137803"
      /translation="MOKYALHAPYVMAIMVATLTGCAWTPAKPIVQAGTTPAQPIPGV
PVANGSIPOSQPINIGYQLFEDRPENIGDTLITVLQENVSASKSSNAASRDGKT
SPGFDTPRYLQGLFGNSRADMGASGSGSPFGKGAASNTSGTLTIVYDOVLANGN
LHVVGKQIAINQCTEFIRFSGVNPRTISGNSVSPSTQVADARIEVYNGYINEAQN
MGMLRFFPLNSPM"
      complement(6350..7132)
      /gene="figc"
      /locus_tag="t1741"
      complement(6350..7132)
      /gene="figc"
      /locus_tag="t1741"
      /note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
      /codon_start=1
      /transl_table=11
      /product="flagellar basal-body rod protein figc"
      /protein_id="AA069365.1"
      /db_xref="GI:29137804"
      /translation="MISLWIAKTGLDRAQNTMIVIANNLAVSTNGFKRQAVFEDL
LYQTLRQPGASSEQTLPGLQIGTGRPAERLERLSQGLSQTNNSKQVAKGGQF
FOVWLPDGTSAVYRTDGSFOVDQNGQLVAGGFQVQPAITIPANALSTIGDVSVT
QQCGAARVQVQQLTLTFMNDTGLSISGENLYETSSGAPNESTPGLNGAGLLYGY
VETSNVAEELVNIQVORAYEINSAVSTTDMLQKLTQL"
      complement(7146..7901)
      /gene="figf"
      /locus_tag="t1742"
      complement(7146..7901)
      /gene="figf"
      /locus_tag="t1742"
      /note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
      /codon_start=1
      /transl_table=11
      /product="putative flagellar basal-body rod protein figf"
      /protein_id="AA069366.1"
      /db_xref="GI:29137805"
      /translation="WDHAIYTMGASQTLNQCAVITASNLNASTPGFRAQNALRAV
PVDLSLATRLTLVASTPGADMTQGLDYTSRPLVDVALQDDGMLVVOAAGAEQYTN
GNIOVPTGQLTIGSHFVIGGSPITVPEGSEITADGITSALNPDDPNTVAPVR
LKLYAKGNEVQRSDGDLFRLLTAQAERGAERGLVLAADSRIMSGVLSNVKPEAVMT
DMIANARFEMQMKVITSVDNEGRANQLLSMS"
      complement(7922..9133)
      /gene="figB"
      /locus_tag="t1743"
      complement(7922..9133)
      /gene="figB"
      /locus_tag="t1743"
      /note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
      /codon_start=1
      /transl_table=11
      /product="flagellar hook protein figB"
      /protein_id="AA069367.1"
      /db_xref="GI:29137806"
      /translation="MSEFQAVSGCLAAATNLDVGNRIANSATYGFKSGTASPADMA
GSKVGLGVKAGITQDFTDGTITNGDLVALSNGFRLVDSNGSVFYSRNGQFCL
```

Query Match 86.7%; Score 395.2; DB 1; Length 301983;
Best Local Similarity 91.7%; Pred. No. 9.8e-103;

```
Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1 ATGAACTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTGTGGCAGTGTCTGCT 60
Db 37210 ATGAACTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTGTGGCAGTGTCTGCT 37251
QY 61 GGGCGTCTTCAAAATGGGGGGGGGGGGTAAATATAGTAAAGGGGGGCAATAGTTCCGGC 120
Db 37250 GGGCGTCTTCAAAATGGGGGGGGGGGGTAAATATAGTAAAGGGGGGCAATAGTTCCGGC 37191
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAAGCGTGGCGTCTCTGCA 180
Db 37190 CCGGATTCACGTTGAGCATTTATCAGTACGGTTCCGCTAAGCGTGGCGTCTCTGCA 37131
QY 181 AGCGATCCCGGTAATCTGAAACGACCATTAACCAAGCGGTTATGTGTAACGGCGCGAT 240
Db 37130 AGCGATCCCGGTAATCTGAAACGACCATTAACCAAGCGGTTATGTGTAACGGCGCGAT 37071
QY 241 GTAGCGAGGGTGGGTAATAGTACTATTGAATGAACTGACCTGAGATGGTTTCAGAAATAT 300
Db 37070 GTAGCGAGGGTGGGTAATAGTACTATTGAATGAACTGACCTGAGATGGTTTCAGAAATAT 37011
QY 301 GCCACCATGACACGATGGAACGCTAAAACTCCGATATTACTGTGGCCATATGATGAC 360
Db 37010 GCCACCATGACACGATGGAACGCTAAAACTCCGATATTACTGTGGCCATATGATGAC 36951
QY 361 CTGGTTACCCGTTGTTTACCATGAAATGGACATGCAAGCGTATGTGCTGAGGTT 420
Db 36950 AATTAACCCCGCTGCTGTTATCAGACCGCATCTGATCCAGCGTAATGTGCTGAGGTT 36891
QY 421 GGTTTGGCAACAGCGCAGCGGCTAACCATATTTAA 456
Db 36890 GGTTTGGCAACAGCGCAGCGGCTAACCATATTTAA 36855

RESULT 7
STAGBPA
LOCUS STAGBPA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AF000514
VERSION AF000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1
REFERENCE
AUTHORS Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
TITLE Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE
AUTHORS Sukupolvi,S.S.
TITLE Direct Submersion
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinamyllykatu, 20520, FINLAND
FEATURES
SOURCE
1..1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
14..553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="agfB protein"
/protein_id="CA04150.1"
```



```

/db_xref="GI:2275120"
/translation="MKKILFLPMLTIIIEPIATYVDLARSSTYFAVNEIKSSPNO
AIIQVGTDSARVRHSGSLISIEBGGRNRRKADVDGYNFAYIEQGNANDA
SISQAYGNSAIIISAIIOKSGSKNITOTYGTOKTAAVVOKSHMAIOANTYGTG
KTAVVOKOSHMAIIVTOR"
593..1048
/gene="agfa"
/cds_start=593
/cds_end=1048
/gene="agfa"
/codon_start=1
/codon_end=1048
/product="Agfa protein"
/protein_id="CA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:O33802"
/translation="MKTLVAAFAAIVVSGSVAAGVVPOMGGGNNHNGSSGPDST
LSIYOGSANAALQSDARKSETTITSGYNGADVGGADNSTITLTONFRNNAT
IDPMNAKNSDITVGYGNNALVNOQASDSSVVRQYGRNNANPANQIN"

```

Query Match
Best Local Similarity 85.4%; Score 389.4; DB 1; Length 1048;
Matches 414; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTGTCTGCT 60
DB 593 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTGTCTGCT 60
QY 61 GGGCTGCTTCCAAATGGGGGGGGGGGGTATATATACGGCGCGCAATAGTTCCGGC 120
DB 653 GGGCTGCTTCCAAATGGGGGGGGGGGGTATATATACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTCAAGTGTAGATTTATACGATCGTTCCGCTAACGCTGCGCTCTGCA 180
DB 713 CCGGACTCAAGTGTAGATTTATACGATCGTTCCGCTAACGCTGCGCTCTGCA 180
QY 713 CCGGACTCAAGTGTAGATTTATACGATCGTTCCGCTAACGCTGCGCTCTGCA 180
DB 181 AGGATGCCCCGTAATTTGTAACGACCATTAACGAGGGGTATGTAACGGCGCAT 240
QY 773 AGGATGCCCCGTAATTTGTAACGACCATTAACGAGGGGTATGTAACGGCGCAT 240
DB 773 AGGATGCCCCGTAATTTGTAACGACCATTAACGAGGGGTATGTAACGGCGCAT 240
QY 241 GTAGCGGAGGTGGGTAATAGTACTATGTAAGTCAATGTTTGAATAAT 300
DB 833 GTAGCGGAGGTGGGTAATAGTACTATGTAAGTCAATGTTTGAATAAT 300
QY 833 GTAGCGGAGGTGGGTAATAGTACTATGTAAGTCAATGTTTGAATAAT 300
DB 301 GCCACATGACCATGTAAGTCAATGTAATGTAATGTAATGTAATGTAATGTAAT 360
QY 893 GCCACATGACCATGTAAGTCAATGTAATGTAATGTAATGTAATGTAATGTAAT 360
DB 893 GCCACATGACCATGTAAGTCAATGTAATGTAATGTAATGTAATGTAATGTAAT 360
QY 361 CTGGTACCCGCTGTTTACCGCAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
DB 953 AATAAGCGCGGCTGTTTACCGCAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
QY 421 GGTGGGCAACAGCCGCGCTAACGAGTATTA 455
DB 1013 GGTGGGCAACAGCCGCGCTAACGAGTATTA 455

```

RESULT 8
LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 56 from patent US 5635617.
ACCESSION 144908
VERSION 144908.1 GI:2469621
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 361)
AUTHORS Doran, J.L., Kay, W.W., Collinson, S. Karen, and Clouthier, S.C.
TITLES Methods and compositions comprising the agfa gene for detection of
JOURNAL Patent: US 5635617-A 56 03-JUN-1997;
Salmonella

FEATURES
source Location/Qualifiers
1..361
/organism="unknown"
/mol_type="unassigned DNA"

Query Match
Best Local Similarity 53.0%; Score 287.2; DB 6; Length 361;
Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 64 GTCTTCCAAATGGGGGGGGGGGGTATATATACGAGCGGTTATGTAACGGCGCATGTA 243
DB 1 GTCTTCCAAATGGGGGGGGGGGGTATATATACGAGCGGTTATGTAACGGCGCATGTA 243
QY 124 GACTCAACGTTGAGCATTATCATGACGTTCCGTAACGCTGCTGCTCGCAAGC 183
DB 61 GACTCAACGTTGAGCATTATCATGACGTTCCGTAACGCTGCTGCTCGCAAGC 183
QY 184 GATCCCGTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCATGTA 180
DB 121 GATCCCGTAAATCTGAAACGACCATTAACGAGCGGTTATGTAACGGCGCATGTA 180
QY 244 GGGCAGGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 303
DB 181 GGGCAGGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 303
QY 304 ACATGACACAGTGGAGCGTAAACCTCCGATTTACTGTCGGCCAAATG 355
DB 241 ACATGACACAGTGGAGCGTAAACCTCCGATTTACTGTCGGCCAAATG 355

```

RESULT 9
LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter sp. Fec2 csb gene, csbA gene and csbD gene.
ACCESSION AJ515700.1 GI:31790491
VERSION AJ515700.1 GI:31790491
KEYWORDS csbA gene; csbB gene; csbD gene; curlin monomer; regulator protein.
SOURCE Citrobacter sp. Fec2
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Citrobacter.
REFERENCE 1
AUTHORS Zogaj, X., Bokranz, W., Nimtz, M. and Romling, U.
TITLES Production of Cellulose and Curli Fimbriae by Members of the Family
JOURNAL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
ACCESSION Infect. Immun. 72 (7), 4151-4158 (2003)
AUTHORS Romling, U.
TITLES Direct Submision
JOURNAL Submitted (11-NOV-2002) Romling U., Microbiology and Tumourbiology
Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
source Location/Qualifiers
1..2889
/organism="Citrobacter sp. Fec2"
/mol_type="genomic DNA"
/isolate="Fec2"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
/complement(212..862)
/gene="csbD"
/complement(212..862)
/gene="csbD"
/functions="necessary for transcription of the csbA
operon"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"


```

/translation="MFNEFHSIHTLTLITKPSLOATALLQHLKOSLALSGKLNIO
RSLDISSCIVLDMMEADKLIHYQDNLSRKNNIKTLLNTPDYPRIENMP
HINCFVADDERVNGLOGLIRECFPSOKLASYLTHSGNRYNSSEALLTHRE
KEILANKRIGASNEIARSFLISENTVKTLYNLFKLIAYNRTQAVSWANDLRR"
1623..2078
/gene="csgb"
1623..2078
/gene="csgb"
/function="nucleator"
/codon_start=1
/translation="curli subunit"
/codon_start=1
/product="curlin-csga protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790493"
/translation="MKRKLPMMLTVLGPAGIASNTSYDLAHSEYFAVELSKSP
QAALIGOVINNSAMRQESKLSVDSQSGSNRAKVDGCAVPAIAGSHSND
SISQSYNGNTAMIIQKSGNKNANTIGYQKTAVVVQROSMAIRVQR"
2119..2571
/gene="csga"
2119..2571
/gene="csga"
/function="curli subunit"
/codon_start=1
/product="curlin-csga protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLIQVAFPAIIVVSGALAGVPQMGCGGGGSSGPESTL
SIYQGVNNAALALQSDARKSDITIHQFNGADVQGSNDSTIDLQNFKNATTI
DQWNGKNSDITVSOYGGHNAALVNGTASDSVLVHQVFGNNATANGY"

ORIGIN
Query Match 61.8%; Score 281.6; DB 1; Length 2889;
Best Local Similarity 77.6%; Pred. No. 4.3e-70;
Matches 354; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

Qy 1 ATGAAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTTCTGGCAGTCTGCT 60
Db 2119 ATGAAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTTCTGGCAGTCTGCT 2178

Qy 61 GGCCTGTTCCACAATGGGGCGCGCGGTATCATMACGGCGCGCATAGTTCCGGC 120
Db 2179 GGTCTGTTCCACAATGGGGCGCGCGGTATCATMACGGCGCGCATAGTTCCGGC 2235

Qy 121 CCGGACTCAAGCTTGAAGATTTATCAATGAGTTCCTCCGTAACGCTGCTCTGCA 180
Db 2236 CCGGACTCAAGCTTGAAGATTTATCAATGAGTTCCTCCGTAACGCTGCTCTGCA 2295

Qy 181 AGCGATGCGGTAAATCTGAAGACGATTCACGAGCGGTATAGTAAACGCGCGAT 240
Db 2296 AGCGATGCGGTAAATCTGAAGACGATTCACGAGCGGTATAGTAAACGCGCGAT 2355

Qy 241 GTAGGCGAGGTGCGGATTAATAGTAACTGATCAAGATGTTTCAAAATAT 300
Db 2356 GTAGGCGAGGTGCGGATTAATAGTAACTGATCAAGATGTTTCAAAATAT 2415

Qy 301 GCCACATCGACCAATGGAAGCTTCAATATTAATGCTGCGCAATATGATGAG 360
Db 2416 GCCACATCGACCAATGGAAGCTTCAATATTAATGCTGCGCAATATGATGAG 2475

Qy 361 CTGGTTACCGGTGTTGTTACCATGAATGAGCAATGCAAGCGTAATGTCGTCAG 420
Db 2476 CTGGTTACCGGTGTTGTTACCATGAATGAGCAATGCAAGCGTAATGTCGTCAG 2535

Qy 421 GGTITTTGGCAACAACGCGCGGTAAACGAGTATTA 456
Db 2536 GGTITTTGGCAACAACGCGCGGTAAACGAGTATTA 2571

RESULT 10
CFR515701 2920 bp DNA linear BCT 24-JUN-2003
LOCUS Cfr515701
DEFINITION Cfr515701 csgB gene, csgA gene and csgD gene.

```

```

ACCESSION AU515701
VERSION AU515701.1
KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE Ctrbacter freundii
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Ctrbacter.
REFERENCE 1
AUTHORS Zoga J, X., Bokranz W., Nimitz M., and Romling U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
JOURNAL Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE 2
AUTHORS Romling U.
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
SOURCE 1..2920
/organism="Ctrbacter freundii"
/mol_type="genomic DNA"
/isolate="fec4"
/isolation_source="faeces"
/species="Homo sapiens"
/db_xref="taxon:56"
/country="Germany"
/complement(211..861)
/gene="csgd"
/complement(211..861)
/function="necessary for transcription of the csgBA operon"
/codon_start=1
/translation="putative regulatory protein"
/protein_id="CAD56673.1"
/db_xref="GI:31790496"
/translation="MFNEFHSIHTLTLITKPSLOATALLQHLKOSLALSGKLNIO
RSLDISSCIVLDMMEADKLIHYQDNLSRKNNIKTLLNTPDYPRIENMP
HINCFVADDERVNGLOGLIRECFPSOKLASYLTHSGNRYNSSEALLTHRE
KEILANKRIGASNEIARSFLISENTVKTLYNLFKLIAYNRTQAVSWANDLRR"
1624..2082
/gene="csgb"
1624..2082
/gene="csgb"
/function="nucleator"
/codon_start=1
/translation="curli subunit"
/product="nucleation component of curlin monomers"
/protein_id="CAD56674.1"
/db_xref="GI:31790497"
/translation="MKRKLPMMLTVLGPAGIASNTSYDLAHSEYFAVELSKSP
QAALIGOVINNSAMRQESKLSVDSQSGSNRAKVDGCAVPAIAGSHSND
SISQSYNGNTAMIIQKSGNKNANTIGYQKTAVVVQROSMAIRVQR"
2123..2572
/gene="csga"
2123..2572
/gene="csga"
/function="curli subunit"
/codon_start=1
/translation="curli subunit"
/product="curlin-csga protein"
/protein_id="CAD56675.1"
/db_xref="GI:31790498"
/translation="MKLIQVAFPAIIVVSGALAGVPQMGCGGGGSSGPESTL
SIYQGVNNAALALQSDARKSDITIHQFNGADVQGSNDSTIDLQNFKNATTI
DQWNGKNSDITVSOYGGHNAALVNGTASDSVLVHQVFGNNATANGY"
ORIGIN
Query Match 50.3%; Score 229.2; DB 1; Length 2920;
Best Local Similarity 70.6%; Pred. No. 5.9e-55;
Matches 322; Conservative 0; Mismatches 128; Indels 6; Gaps 1;

```

ORIGIN	Query Match	50.2%	Score 229,	DB 1;	Length 230;
	Best Local Similarity	99.6%;	Pred. No. 5.3e-55;		
	Matches 229;	Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
QY	88	GGTAATCATTAACGGCGGGCGGCAGATAGTTCGCGCCCGGACATCACTGTGACATTATACG	147		
Db	1	GGTAATCATTAANAGCGGGCGGCAGATAGTTCGGCGCCCGGACATCACTGTGACATTATACG	60		
QY	148	TACGGTTCGGCTAAACGCTGCGCTTGCTGTGCGAAGCGATGCCCGTAAATCTGAAAACGACC	207		
Db	61	TACGGTTCGGCTAAACGCTGCGCTTGCTGTGCGAAGCGATGCCCGTAAATCTGAAAACGACC	120		
QY	208	ATTACCCAGAGCGGTTATGTGTAAACGGCGCCGATGTATGAGCGAGGTTGCGGATATATGTA	267		
Db	121	ATTACCCAGAGCGGTTATGTGTAAACGGCGCCGATGTATGAGCGAGGTTGCGGATATATGTA	180		
QY	268	ATTGAACAGACTCAGAAATGGTTTCAGAAATTAATGCCACATGACACCAATG	317		
Db	181	ATTGAACAGACTCAGAAATGGTTTCAGAAATTAATGCCACATGACACCAATG	230		

RESULT 12	
ECCSGABDG	
LOCUS	4680 bp
DEFINITION	E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION	X90754
VERSION	X90754.1
KEYWORDS	csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene; orfC gene.
SOURCE	Escherichia coli
ORGANISM	Escherichia coli
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE	1
AUTHORS	Hammar, M., Arngvist, A., Bian, Z., Olsen, A. and Normark, S.
TITLE	Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia coli K-12
JOURNAL	Mol. Microbiol. 18 (4), 661-670 (1995)
MEDLINE	96414468
PUBMED	8817489
REFERENCE	2 (bases 1 to 4680)
AUTHORS	Hammar, M.
TITLE	Direct Submission
JOURNAL	Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm, SWEDEN
FEATURES	location/Qualifiers

```

URS
source
location/Qualifiers
1. .4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/map="23.15 minutes"
1. .3140
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="WC4100"
/db_xref="taxon:562"
3141. .4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="W3110"
/db_xref="taxon:562"
repeat_region
<1. .>4680
/insertion_seq="IS2 (partial)"
complement(133. .966)
/gene="csaG"
complement(133. .966)
/gene="csaG"
CDS
/codon_start=1

```

```

/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CA62277.1"
/db_xref="GI:1147559"
/db_xref="GOA:P52103"
/db_xref="SWISS-PROT:P52103"
/translation="MORFLVAVMLSGCLTAPKREARPTLMPRAQSYKDLTHLPA
PYGKIFVSVVNIQDETQPKFPASNPSTANPQSATMLVYALKDSRFITLBERQGLQ
NLNBRKTIKRAQENGTAVNNRIPLQSLTANIMVBSIIIGESNVKSGGVARFEG
IGADTQYOLDQIAVLRVNVSTGEIISVNTSKTILSYEQAGVFRIIDYQRLLEG
VGYSNBEVMLCLMSALETGVIFLINIDIDRGLMDLQNKARQNDILVYKSHMSVPE
S"
complement(993..1409)
/gene="csge"
complement(993..1409)
/gene="csge"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CA62278.1"
/db_xref="GI:1147560"
/db_xref="GOA:P52104"
/translation="MRVHAVVLLMLISPLSMAGTMTFOFPNPNFGNPNNGAFILNS
AQQNSYKDPYNDPFIETPSALDNFTQAIQSLIGSLSNINMGKRGKRVNTDIYV
DIARBDQLQNLNDRKTCGRSTIQVSGLOKNSIDF"
complement(1434..1823)
/gene="csge"
complement(1434..1823)
/gene="csge"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CA62279.1"
/db_xref="GI:1147561"
/db_xref="GOA:P52105"
/translation="MKRYLRWIVAAEPIFAAGNHAHVEVPEGLTHTYSSIGHDY
RASFDRKESDTGULTINERPSARWGSHITTVQDVIFQTFPLKRDPEKTIVFAL
IQTEALNRRQIQALSTGDLADEF"
complement(1828..2478)
/gene="csge"
complement(1828..2478)
/gene="csge"
/translation="necessary for transcription of the csge operon"
/translation="necessary for transcription of the csge operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
/protein_id="CA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translation="MFNEVSHIGHITLLITKSLQATLALQHLKQSLATIGKLNIO
HSLNDISSGIIILDMBADKLIHYMODTSLRNNNKKILLNTPEDPYRDIENP
RINDGVFMEQDEEVNGLOGVLRGECYFOKLASVLIISGNRVNSTSAITLTHE
KEIANKRIRICSHNNEIARSLPISNTYKTHYMLFKIAYVNRIOAVSWANDIIR"
complement(2626)
/translation="transcriptional start site for csge operon"
/translation="transcriptional start site for csge operon"
complement(3140)
/translation="transcriptional start site for csge operon"
complement(3233..3688)
/gene="csge"
complement(3233..3688)
/gene="csge"
/codon_start=1
/transl_table=1
/product="nucleation component of curli monomers"
/protein_id="CA62281.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translation="MKNKLLFMMLTIIGAPGIAAAGVLANSEYPAVNBLSKSFN

```

```

OAAIIQAGTNNNSAQLRQGSKLLAVVAQESSNRAXIDQTDVNLAYIDQASANDA
SISQAGCNTAMITIQKSKNKANITQYGTQKTAIVQROSMALRVQR"
3729..4184
/gene="csge"
3729..4184
/gene="csge"
/translation="major subunit"
/codon_start=1
/transl_table=1
/product="curliin/Csge protein"
/protein_id="CA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="MKLLKVAIATVIFSGSALAGVVPYQGGGNHGGGNGSPNSE
LNIYQGGNSALALQTDARNSDITTIQHGKGNADVGQSDSSIDLTQRFNSAT
LDQWNGKNSMTVYQFGGNGAADVQTAANSNVVTVQFGNNATAHQY"
3729..3788
/gene="csge"
4243..4575
/gene="orfC"
4243..4575
/gene="orfC"
/codon_start=1
/transl_table=1
/product="putative curli production protein"
/protein_id="CA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="MNTLLILALSSQITENTTQOQGVYTIPEVTLTQSLCRVQIL
SLREGSGQSGQTKQEKTLSPANQPIALTKLSINISPDDBVKIVVTVSDQSLHLSQO
WPSSESS"

```

ORIGIN

Query Match 49.8%; Score 227.2; DB 1; Length 4680;
 Best Local Similarity 68.6%; Pred. No. 2.3e-54;
 Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTGTGGCAGTCTCTGGCT 60
Db 3729 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTGTGGCAGTCTCTGGCA 3788
QY 61 GGGCGTTCACAAATGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
Db 3789 GGTGTTTCTCTCACTACGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 3848
QY 121 CCGGACTCAAGCTTGACATTTATCAGTACGGTTCCGCTAACGCTGCGCTTGTCTGCAA 180
Db 3849 CCAATTCTGAGCTGAACATTTACAGTACGGTGGCGGTAACTTGCACCTTGTCTGCAA 3908
QY 181 AGCGATCCCGTAAATTTGAAACGACATTTACCAAGCGGTTATGTTAAGCGCGCGAT 240
Db 3909 ACTGATCCCGTAACTTGAATTTGACATTTACCAAGATGCGCGGTTATGTTAAGCGAT 3968
QY 241 GTAGCGCAGGAGTGGCATTAATGACTATTTGAAGTGAAGTGAAGTGGTTTCAGAAATAT 300
Db 3969 GTTGTGAGGCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4028
QY 301 GCCACATCGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 360
Db 4029 GCTACTCTTATCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4088
QY 361 CTGGTTACCGGTGTTTGAATCCATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 420
Db 4089 GGCACCGTCTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 4148
QY 421 GGTTCGACACAGCGCCTAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 456
Db 4149 GCGTTGTGAACACAGCGCCTAACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4184

```

RESULT 13
 AE000205

```

LOCUS       AE000205               10346 bp    DNA             linear    BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete
ACCESSION   AE000205 U00096
VERSION     AE000205.1 GI:1787265
KEYWORDS
SOURCE
ORGANISM    Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1 (bases 1 to 10346)
AUTHORS     Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
            Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
            Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
            Mau, B. and Shao, Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5311), 1453-1474 (1997)
TITLE       JOURNAL
MOLTYPE     PUBMED
PUBMED      9278503
REFERENCE   2 (bases 1 to 10346)
AUTHORS     Blattner, F.R.
TITLE       Direct Submission
JOURNAL     Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3 (bases 1 to 10346)
AUTHORS     Blattner, F.R.
TITLE       Direct Submission
JOURNAL     Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4 (bases 1 to 10346)
AUTHORS     Plunkett, G. III.
TITLE       Direct Submission
JOURNAL     Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using Genemark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the database is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M55. No
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             location/Qualifiers
     source           1..10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
                     complement(1..48)
     repeat_region    complement(1..48)
     gene             1..5
                     /note="IS3"
                     /note="factor Sigma70; predicted +1 start at 1094689"
                     /note="factor Sigma70; predicted +1 start at 1094716"
                     /gene="b1028"
                     /gene="b1029"
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="orf; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1888 aa protein Fas2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYIAGLDIASPEHGVIANVTAKSEVAPSIKENTLS
                     LHLTHVKNYTESLDITFLNMPFVAFLEISNKPYLGNMINDPAGVDQDQVHRCRAFSQ
                     AKYFK"
                     275..299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
                     /gene="ycdu"
                     /note="synonym: b1029"
                     /gene="ycdu"
                     /note="synonym: b1029"
                     /gene="ycdu"
                     /function="orf; Unknown"
                     /note="orf; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCB
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLFIRFODKRLIYIATGILIGPYWKNAGPTPSEDI
                     GVSGIALVYVNFIPDLKAYKAYCTKNDSPKKNHIEFLTPDVAFLS
                     LIMASMSWGLYOLPSLVAFLISLGLVITFLRMRTSYKQVALSAKAVKX
                     SLTRYVILSVGISVNVLLTISPRNDSFTVGQMTKRSITIALILGCVLAINLF
                     PLRSKRYAPLGRFLFDIDLPFSSEALSTFPAKPLMRLFTILVIEVMITLVSVL
                     ATIVEMKWEAFALCYVPLIYFYFCRFLMNDPMACDWFYFMGHFNK"
                     1253..1280
                     /gene="ycdu"
                     /note="central position to predicted promoter: -104"
                     /bound_moiety="Deor predicted site"
                     1253..1280
                     /gene="ycdu"
                     /note="No predicted promoter"
                     /bound_moiety="Deor predicted site"
                     1337..1365
                     /gene="ycdu"
                     /note="factor Sigma70; predicted +1 start at 1096049"
                     1494..1745
                     /gene="b1030"
                     1494..1745
                     /gene="b1030"
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="orf; This 83 aa ORF is 25 pct identical (7 gaps)
                     to 81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="WGGTRCVQYTKKKVLSYELFFYGGEGGIDSLRSCGPANHC
                     VHGLSNMLSPVVDPRSGLLIPVCAIHKKARTVRALL"
                     1513..1926
                     /gene="ycdu"
                     /note="synonym: b1031"

```

```

CDs
1513..1926
/gene="ycdV"
/function="putative structure; Not classified"
/note="O137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein RS37_YEAST SM:
P05759"
/codon_start=1
/transl_table=1
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKSPYFRTSSILMAVRGGLTFAPAPGSLPTAFVCP
GROLSTPGRGSPPRCVQYKKKPVLSYELFFKGGGGIDSLRSPGGAGHCVHGL
SNMLSPVEPRSGLLIPVCAIYKAKRTFVALL"
/complement(2111..2198)
/gene="serX"
/note="synonym: b1032"
/complement(2111..2198)
/gene="serX"
/product="tRNA-Ser"
/note="anticodon: CGA; CG Site No. 17878"
/function="tRNA; tRNA"
2269..2296
/note="factor Sigma70; predicted +1 start at 1096980"
2338..2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355..2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393..3370
/gene="ycdW"
/note="synonym: b1033"
2393..3370
/gene="ycdW"
/function="putative enzyme; Not classified"
/note="O325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPRA_CORGL
SM: P45637"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="MRINFAORFSGVSMIDIFYPPTDTQWMEALRKAIPOARVAR
KSGDNDADVALVWHPVEMLAGRLKAVPALGVDUSLSTKQHPKEMNPVLR
LEDTGSEOMOEIVASOVLHFRFRPDYRIQNSHNPDPYHREDFTTIGLAGVL
GSKVASLQWTRPRLCWSRTRKSPGVQSPAGEELSAFISQCRVILNLPNPEV
GIINQQLLEKLPGDAVILNLAGVHVEDLLALDSGKVGAMLDVFNHEPLPESP
LMQHPRTIIPVHAITRPAVAEYISRTIAQLEKGERVCGVDPRARGY"
3278..3308
/gene="ycdW"
/note="factor Sigma70; predicted +1 start at 1097992"
3383..3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396..3423
/promoter

Query Match 49.8%; Score 227.2; DB 1; Length 10346;
Best Local Similarity 68.6%; Pred. No. 2.5e-54;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Oy 1 ATGAAATCTTTTAAAGTGCAGATTCGACAGCATTCGAGTTCTGCGAGTCTTGAGCT 60
Db 8993 ATGAAATCTTTTAAAGTGCAGATTCGACAGCATTCGATTCCTCGGAGGCTCTGGCA 9052

Oy 61 GGCCTGTTCCACATGGGGCGCGCGGATATATTAACGGCGCGGCAATGTTCCGGC 120
Db 9053 GGTCTTCTCTCAGTACGGGGCGCGGATTAACCGGTGCGGATTAATAAGCGGC 9112

Oy 121 CCGGACTCAACGTTGAGATTATCATAGAGGTTCCGCTAACGGTGGCTTGCTCTGCA 180
Db 9113 CCAAAATCTGAGCTGAACATTTACAGTACGTTGCGGATCTTGCACTTGCTCTGCA 9172

Oy 181 AGCGATGCCGTAATCTGAACAACGACATTACCGAGACGGTTATGGTAACGGCGCCGAT 240

```

```

Db 9173 ACTGATGCCGTAATCTGACTGACTTATTAACGACATGGCGGCGTATGTGCAGAT 9232
Oy 241 GTAGCCAGGGTGGGATATATGACTTATTAAGTCTGAAATGTTTCAAAATAT 300
Db 9233 GTTGTAGGGCTCAGATGACAGCTCAATCATCTGACCCAACTGGCTCGTTAACAGC 9292
Oy 301 GCCACATCGACGATGGGAACGTAAAAACGCCGATTTATCTGCGGCATATGATCAG 360
Db 9293 GCTACTCTGATCATGTGAACGCGAANAATTTCTAAATGACGGTTAAACATTCGCTGCT 9352
Oy 361 CTGTTACCCGTTGTTTACCATGAATAATGACATGCAATGCAATGATGTCGTCAGGTT 420
Db 9353 GGCACAGGTGTGAGATTGACACAGTCAGCTAATCTCTCCGTCACATGATCAGGTT 9412
Oy 421 GGTTTGGCAACAGCCAGCGCTAACCAATTTA 456
Db 9413 GGTCTTGTTAAACAGCGACCGCTCATCAGTACTA 9448

RESULT 14
D90741
LOCUS
DEFINITION
D90741 AB001340
ACCESSION
D90741.1 GI:1651509
VERSION
D90741.1 GI:1651509
KEYWORDS
Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
csgA; ycdB; c1s; nov; mdoG.
SOURCE
Escherichia coli K12
Organism
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Oshima,T., Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nishio,H.,
Nishio,Y., Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K.,
Wada,C., Yamamoto,Y., Yano,M. and Horichi,T.
A 718-Kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL
MEDLINE
97061202
PUBMED
8905232
2
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horichi,T.,
Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
Motomura,K., Nakamura,Y., Nishio,H., Nishio,Y., Oshima,T.,
Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
Yamamoto,Y. and Yano,M.
The systematic sequencing of the Escherichia coli genome in Japan
Unpublished
3 (bases 1 to 15047)
Mori,H.
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Direct Submission
Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
and Technology, Res. & Edu. Center for Genetic Info., 8916-5
Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@etc.aisf-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669)
Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
Horichi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
Nakamura,Y., Nishio,H., Nishio,Y., Oshima,T., Saito,N.,

```

Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.
 Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishori@nibb.ac.jp
 Information operator:
 Name: Hirotsada Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmori@gsc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://bwsj.aist-nara.ac.jp
 Location/Qualifiers
 1..15047
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:83333"
 /map="23.7-24.0 min"
 /clone="Kohara clone #231"
 /note="Nucleotide position 1097101-1112147 from the initiation site of Theta (0 min.). This clone is from Kohara lambda miniset library."
 <1..276

gene

CDS

CDS

CDS

CDS

CDS

CDS

```

/note="unnamed protein product; ORF_ID:0230#5"
/codon_start=1
/transl_table=11
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translacion="DLFPSSSENALSTFPAPKPLMLRFLILVIEVMITLVSALATIVE
WRIFWFAFLLCYVPCLIYFFYCRFLMNDPMACDMYFRMGHENK"
1152..2090
/note="ORF_ID:0231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=11
/product="Hypothetical 33.0 kd protein in prob-prok
intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translacion="MDIIFHPPTDQWMEALRKAIPQARVAMKSGNDSADVALV
WHPVEMLAGRDILKAVAGVDSILSKQAPENMLPSVPLFLEDQMGOMQOEY
AVSQVLFRRPDYRQONSSHOMPLPEYHRDPFLITIGAGVIGSKVAOSICTMRF
PLRCWSTRKSPGVOSFAGRELSAFLISOCRVLTINLPTPTVGIINQOLLEKIPD
GAYLINLRGVHVEDDILALDSCKYKGMVLVPRBPLPBPSPLMQHRVTTTHV
AATTPRAVEYISRTTIAQLKGERVCGQVDRARGV"
2145..2882
/note="unnamed protein product; ORF_ID:0231#2"
/codon_start=1
/transl_table=11
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translacion="WTFVDLHMTVASTAVSTLSDYIAQAKQKIKLPAITDGHDM
EDBPHMHFTNMRIMPRVVDGIIIGIEANIKNDGVIDCSGKMFDSIDLIIAGFHE
VPAPHDKATNTQAMATIAASGNVHIISHGPNKXYIDYKAAVEAAKXKQVLEINNS
SPHRSKGSSEDCREVAALAVRDAGWVALGSDSHTAFTNGEPBECIKILDADVDPPEP
ILVSPRRLNLFESRGMAPIAEFADL"
2906..3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein H11543"
/db_xref="GI:4062608"
/translacion="MNEFSLICRVGLSVYROPQDPLVLPFTLLREKGLANMPLQ
DPTLRLQSCMTQVSADYNALFTIGDECAVPYPSAWEGATEAVEAFISERGMPL
AATPDHITGLTLAASWMLDOSTEDSEALETLFSEYLLPMCGAFIIGKVEAATTPFW
RTMAPLTRDAISAMWDELDEDSSE"
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"

```

gene

CDS

CDS

gene

CDS

gene

gene

```

/codon_start=1
/transl_table=11
/protein_id="BAA35826.1"
/db_xref="GI:4062609"
/translacion="MCSISFMAFAFAIMKGNMILSIATTTGILSGIMGWAVSLGL
SMAGFLGCTAVFACPOGGKGLAISATILSGVWMMVVIYSGALAPHEIIGVYTG
IVAFMNCIOAKQOLLSFVPDFTGACATPAGODMVLVPSIALGLIFGAMKNSGLM
LAARSAKTAREBOEIKNKA"
complement (4117..4950)
/gene="cs9g"
complement (4117..4950)
/gene="cs9g"
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
cs9g precursor."
/protein_id="BAA35827.1"
/db_xref="GI:4062610"
/translacion="MORLEFLVAVMLLGGCLTAPPKBAAPPTLMPRAQSYKDLTHLP
PGKTFVSYVNIODEGOFKPYPASNFSTAVPOSATMLYTALKSRPFLPERGLO
NLINRKTIIRABOEGTVAIINRILQSTITANIVGESITGESNVSGGVGARFG
ICADYQYQDQJAVNLVRVNVSTGELISVNTSKITLSIEVQAGVFRFLDQRLLEG
VGTNSNEPMLCLMSAIEGTVEFLINDGIDRLMDLONKAERONDILVKYRMSVPE
S"
complement (4977..5393)
/gene="cs9f"
complement (4977..5393)
/gene="cs9f"
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
cs9f precursor."
/protein_id="BAA35828.1"
/db_xref="GI:4062611"
/translacion="MRVHAAVLLMLISPLSWAGTMTFQFPNPNFGNPNNGAFLINS
AQONSYKDPNDPFGIETPSALDNFTQALQSOILGLGLSINITGKRGKRVNTDYIV
DIANDQLOLNVDRTKGTGTIOVSLQNNSTDF"
complement (5418..5807)
/gene="cs9b"
complement (5418..5807)
/gene="cs9b"
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
cs9b precursor."
/protein_id="BAA35829.1"
/db_xref="GI:4062612"
/translacion="MKRYLRWIVAAEFLFAAGNLHAAVEVPGLLTDHTVSSIGHDPY
RAFSDKWSDETGNLTINERSAPGSMITTVVQDVIFQTFPLPRRDEKTYVFL
IOTEBALNRDIONALSTGDLADER"
complement (5812..6462)
/gene="cs9d"
complement (5812..6462)
/gene="cs9d"
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=11
/product="Probable cs9b operon transcriptional regulatory
protein."
/protein_id="BAA35830.1"
/db_xref="GI:4062613"
/translacion="MEWESIHGHTLLITKSSLOATALLQHKOSIATGKQIANQ
RSJDDISGSGIITLDDKADKLIHYQDTLSKNNIKLILNLTDPDIPYRIENBP
HINQVYSMEDQERVNGLOGVARGECYFOKLASVLIHISGVYRNSTSSALLTHEE
KEILNKIRIGASNNELARSLPISENTVYKHLVNLFKIAKNTQAVSWANDMLRR"
7217..7672

```

CDS /gene="csb"
7217..7672

Query Match 49.8%; Score 227.2; DB 1; Length 15047;
Best Local Similarity 68.6%; Pred. No. 2.6e-54;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
OY 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGATGTTCTGGAGTGGCTGCT 60
DB 7713 ATGAACTTTTAAAGTGGACGATTCGACGATCGATGTTCTGGAGTGGCTG 7772
OY 61 GCGCTGTTTCCACATGGGGGCGCGCGGTATCATTAACGGCGCGCATAGTCCG 120
DB 7773 GGTGTTGTTCTCAGTACGGCGCGCGGTATCATTAACGGCGCGCATAGTCCG 7832
OY 121 CCGACATCAAGCTTGAAGATTTATCAATGAGCTTCCGTTACGCTGCTCTGCA 180
DB 7833 CCAAAATCTGAGCTGAACATTTACAGTACGATGCGGTAACTGCACTTGCTGCA 7892
OY 181 AGCATGCGCGTAAATCTGAACGACATTAACGACGCGTTATGTAACGGCGCG 240
DB 7893 ACTATGCGCGTAACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 7952
OY 241 GTAGCCAGGCTGGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
DB 7953 GTTGTAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 8012
OY 301 GCCACATCGACGATGGAAGCTTAAATCTCCATATTAATGATGATGATGAT 360
DB 8013 GCTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 8072
OY 361 CTGTTACCGCTGTTGTTTACCATGAATGGACATGCAAGGTAATGATGATG 420
DB 8073 GGCACGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 8132
OY 421 GGTTTGGCAACACGCGACGCTAACCGATTTAA 456
DB 8133 GGTGTTGTAACACGCGACGCTCATCATGATCTAA 8168
```

RESULT 15
AX814811 456 bp DNA linear PAT 05-DEC-2003
LOCUS Sequence 15 from Patent WO03064446.
DEFINITION AX814811
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS Bioerck, L., Olsen, A., Wikstroem, M. and Herwald, H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
FEATURES
source
1..456
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..456
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB83541.1"
/db_xref="GI:39104002"
/translation="MKLVKVAIPAIVFSGSALAGVVPYGGGNGHGGGNGGPNSE
LNIYVGGSALALQTDARNSDLITTOHGGGNGADYGGGSDSSIDLTOHGFNSKAT
LDMNGNSKNTVAKQFGGNGAVALDQASNSVAVTQVGFNNATAHQY"

ORIGIN

Query Match 49.5%; Score 225.6; DB 6; Length 456;

Best Local Similarity 68.4%; Pred. No. 5.5e-54;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

```
OY 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGATGTTCTGGAGTGGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGATGTTCTGGAGTGGCTGCTG 60
OY 61 GCGCTGTTTCCACATGGGGGCGCGCGGTATCATTAACGGCGCGCATAGTCCG 120
DB 61 GGTGTTGTTCTCAGTACGGCGCGCGGTATCATTAACGGCGCGCATAGTCCG 120
OY 121 CCGACATCAAGCTTGAAGATTTATCAATGAGCTTCCGTTACGCTGCTCTGCA 180
DB 121 CCAAAATCTGAGCTGAACATTTACAGTACGATGCGGTAACTGCACTTGCTGCA 180
OY 181 AGCATGCGCGTAAATCTGAACGACATTAACGACGCGTTATGTAACGGCGCG 240
DB 181 ACTATGCGCGTAACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 240
OY 241 GTAGCCAGGCTGGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 GTTGTAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATG 300
OY 301 GCCACATCGACGATGGAAGCTTAAATCTCCATATTAATGATGATGATGAT 360
DB 301 GCTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
OY 361 CTGTTACCGCTGTTGTTTACCATGAATGGACATGCAAGGTAATGATGATG 420
DB 361 GGCACGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 420
OY 421 GGTTTGGCAACACGCGACGCTAACCGATTTAA 456
DB 421 GGTGTTGTAACACGCGACGCTCATCATGATCTAA 456
```

Search completed: March 15, 2004, 22:50:05
Job time : 1965.17 secs

THIS PAGE BLANK (USPTO)

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Bacteroides* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match Best Local Similarity 100.0%; Score 456; DB 3; Length 456;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
QY 61 GGGCTGCTTCCCAATGGGCGCGCGCGGTATCATTAACGGCGCGCAATAGTCCGCG 120
DB 61 GGGCTGCTTCCCAATGGGCGCGCGCGGTATCATTAACGGCGCGCAATAGTCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAACGCTGCTCTGGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAACGCTGCTCTGGCA 180
QY 181 AGCGATGCCGTAATTCGAAACGACATTAACCGAGCGGTTATGTTACGCGCGCAT 240
DB 181 AGCGATGCCGTAATTCGAAACGACATTAACCGAGCGGTTATGTTACGCGCGCAT 240
QY 241 GTAGGCCAGGCGGTGGAATTAATGTAATTAATTAATTAATTAATTAATTAAT 300
DB 241 GTAGGCCAGGCGGTGGAATTAATGTAATTAATTAATTAATTAATTAATTAAT 300
QY 301 GCCACCATCGACCGAGTGAACGCTAAATCTCCATTAATTAATTAATTAATTAAT 360
DB 301 GCCACCATCGACCGAGTGAACGCTAAATCTCCATTAATTAATTAATTAATTAAT 360
QY 361 CTGGTTACCGCGTGTGTTATCCATGAATGACATGCAAGCGGTATGTCAGGTT 420
DB 361 CTGGTTACCGCGTGTGTTATCCATGAATGACATGCAAGCGGTATGTCAGGTT 420
QY 421 GGTGTTGGCAACAACGCGCGGTAAACGATTTAA 456
DB 421 GGTGTTGGCAACAACGCGCGGTAAACGATTTAA 456

RESULT 2
AA087467
ID AA087467 standard; DNA; 456 BP.
XX
XX AA087467;
AC
XX 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
XX AgfA sequence.
DE
XX
XX
XX *Salmonella*; AgfA; vaccine; genetic immunization; ds.
OS *Salmonella*.
XX

FH Key Location/Qualifiers
FT CDS 1..454
FT /*tag= a
FT /note= "AgfA"
XX
XX W09425598-A2.
XX
XX 10-NOV-1994.
XX
XX 26-APR-1994; 94MO-IB000207.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UUYI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX (KING/) KING J.
XX
XX Kay WW, Collinson SK, Clouthier SC, Doran JL,
XX WPI; 1994-358275/44.
XX
XX F-PSDB; AAR74625.
XX
XX Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
XX strains, vector constructs, or compans. contg. fimbrial type proteins.
XX Disclosure; Fig 7B; 95pb; English.
XX
XX The DNA encodes the *Salmonella* AgfA protein. The DNA and isolated
XX proteins are used in genetic immunization and vaccine compositions,
XX respectively, to elicit an immune response to *Salmonella* in animals (e.g.
XX food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
XX field.)
XX
XX Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match Best Local Similarity 88.8%; Score 404.8; DB 2; Length 456;

Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
QY 61 GGGCTGCTTCCCAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTCCGCG 120
DB 61 GGGCTGCTTCCCAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAACGCTGCTCTGGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAACGCTGCTCTGGCA 180
QY 181 AGCGATGCCGTAATTCGAAACGACATTAACCGAGCGGTTATGTTACGCGCGCAT 240
DB 181 AGCGATGCCGTAATTCGAAACGACATTAACCGAGCGGTTATGTTACGCGCGCAT 240
QY 241 GTAGGCCAGGCGGTGGAATTAATGTAATTAATTAATTAATTAATTAATTAAT 300
DB 241 GTAGGCCAGGCGGTGGAATTAATGTAATTAATTAATTAATTAATTAATTAAT 300
QY 301 GCCACCATCGACCGAGTGAACGCTAAATCTCCATTAATTAATTAATTAATTAAT 360
DB 301 GCCACCATCGACCGAGTGAACGCTAAATCTCCATTAATTAATTAATTAATTAAT 360
QY 361 CTGGTTACCGCGTGTGTTATCCATGAATGACATGCAAGCGGTATGTCAGGTT 420
DB 361 CTGGTTACCGCGTGTGTTATCCATGAATGACATGCAAGCGGTATGTCAGGTT 420
QY 421 GGTGTTGGCAACAACGCGCGGTAAACGATTTAA 456
DB 421 GGTGTTGGCAACAACGCGCGGTAAACGATTTAA 456

RESULT 3
AAT74142

ID	AAAT74142	Standard; DNA; 456 BP.
XX		
AC	AAT74142;	
DT	25-MAR-2003	(revised)
DT	29-SEP-1997	(first entry)
XX		
DE	Salmonella enteritidis 27655-3b agfa gene.	
XX		
KW	Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.	
OS		
XX	Salmonella enteritidis.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	1..456
FT		/tag= a
FT		/label= agfa gene fragment
FT		/transl_except= (pos:367..369,aa:Pro)
XX		
PN	US5635617-A.	
XX		
PD	03-JUN-1997.	
XX		
PE	26-APR-1994;	94US-00233788.
XX		
PR	26-APR-1993;	93US-00054452.
XX		
PA	(UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.	
PI	Collinson SK, Kay WW, Doran JL;	
XX		
DR	WPI; 1997-309886/28.	
DR	P-PSDB; MAM23570.	
XX		
PT	Isolated Salmonella gene agfa - used for diagnosis of Salmonella or	
PT	enteropathogenic bacteria of the Enterobacteria family.	
XX		
PS	Claim 1; Col 19-112; 85bp; English.	
XX		
CC	The present sequence represents an isolated agfa gene derived from	
CC	Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide	
CC	diagnostic assays for Salmonella and/or enteropathogenic bacteria of the	
CC	family Enterobacteria. It can also be used to provide proteins and	
CC	antibodies which can be used for assays. The nucleic acid sequence can be	
CC	used to provide probes or primers which can specifically hybridise to	
CC	nucleic acid molecules from greater than 99% of Salmonella strains that	
CC	are pathogenic to warm-blooded animals relative to nucleic acid molecules	
CC	from virtually all other microbial organisms. (Updated on 25-MAR-2003 to	
CC	correct PF field.)	
XX		
SQ	Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;	
	Query Match	88.8%; Score 404.8; DB 2; Length 456;
	Beet Local Similarity	93.0%; Pred. No. 1.2e-121;
	Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;	
OY	1 ATGAAGCTTTTAAAAAGTGCGAGCATTCGACGAATCGTAGTTCTGGCAGTGCCTGGCT	60
Dd	1 ATGAAGCTTTTAAAAAGTGCGAGCATTCGACGAATCGTAGTTCTGGCAGTGCCTGGCT	60
OY	61 GGCGTCGTTCACAATGGGGCGGCGCGGTAATCAATACGCGCGGCAATAGTCCGGC	120
Dd	61 GGCGTCGTTCACAATGGGGCGGCGCGGTAATCAATACGCGCGGCAATAGTCCGGC	120
OY	121 CGGAGCTCAAGCTTGAGAGATTATACAGTACGATTCGGTAAACGCTGCGCTGCTGCAA	180
Dd	121 CGGAGCTCAAGCTTGAGAGATTATACAGTACGATTCGGTAAACGCTGCGCTGCTGCAA	180
OY	181 AGCGATGCCCTTAATCTGAAACGACCATTAACCGAGACGGTTATGTAAACGCGCCGAT	240
Dd	181 AGCGATGCCCTTAATCTGAAACGACCATTAACCGAGACGGTTATGTAAACGCGCCGAT	240
OY	241 GTAGCCAGGGTGGGATTAAGTACTAATTAAGTAACTGACTCAAGATGGTTTCAGAAATAAT	300

Db	241	GTAGGCGCAGGGGCGCGAATAATAGTCTATTTGAACGACTCAGAAATGTTTCAGAAAATAAT	3000		
Qy	301	GCCACCATTCGACCACTGGNACGCTAAAAAATCCGATATTACTGTCCGCCAATATGATCAG	366		
Db	301	GCCACCATTCGACCACTGGNACGCTAAAAAATCCGATATTACTGTCCGCCAATACGGCGGT	366		
Qy	361	CTGGTTACCCGCTGTTGTTACCCATGAAATGGAACATGCACGCGTAATGGTGCGTCAGTT	420		
Db	361	AATAACGCGCGCTGTTAATCATCAGACCGCATCTGATTCAGCGTAATGGTGCGTCAGTT	420		
Qy	421	GGTTTTGGCAACACGCGCAGGCTTACCAAGTATTAA	456		
Db	421	GGTTTTGGCAACACGCGCAGGCTTACCAAGTATTAA	456		
RESULT 4					
ID	AAC64617	AAC64617 standard; DNA; 456 BP.			
AC	AAC64617;				
DT	26-FEB-2001	(first entry)			
DE	Salmonella enteritidis Agfa DNA sequence SEQ ID NO:1.				
KV	Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;				
XX	vaccine; immune response; immunogen; ds.				
OS	Salmonella enteritidis.				
PN	WO200060102-A2.				
PD	12-OCT-2000.				
XX	05-APR-2000; 2000WO-CA000356.				
PR	05-APR-1999; 99US-0127888P.				
PA	(UVVI-) UNIV VICTORIA.				
PI	White AP, Doran JL, Collison SK, Kay WW;				
DR	WPI; 2000-672631/65.				
PT	P-PSDB; AAB36341.				
XX	Recombinant agfa gene having a segment replaced by a foreign DNA sequence				
XX	which encodes foreign epitope or antigen, expresses recombinant Agfa				
XX	protein useful for eliciting immune response in animal.				
Disclosure; Page 134; 139pp; English.					
CC	The present invention describes a recombinant agfa gene (1) where a				
CC	segment of the gene has been replaced by a segment of a foreign DNA				
CC	sequence which encodes a foreign epitope or antigen. Also described are:				
CC	(1) use of thin aggregative fimbriae (SFR17/TAF) nucleation depended				
CC	assembly system of strains of Salmonella, Escherichia coli and				
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant				
CC	Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)				
CC	directing recombination of a recombinant gene into the chromosome of the				
CC	homologous species; (3) directing recombination of a recombinant gene				
CC	back into the chromosome of the homologous species, replacing the native				
CC	copy of that gene; and (4) eliciting an immune response in an animal,				
CC	comprising separating an amino acid polymer comprising a recombinant Agfa				
CC	protein containing a replacement segment or segments of foreign amino				
CC	acid sequence or sequences grown on a Salmonella, E. coli or				
CC	Enterobacteriaceae host cell, from the host cell and introducing the				
CC	polymer into the animal in conjunction with a carrier or diluent. (1) is				
CC	useful for the expression of recombinant Agfa protein which is useful for				
CC	eliciting an immune response in an animal. In a fimbrial presentation				
CC	system the heterologous antigens are presented in high numbers (up to				
CC	500,000 copies/cell), the hybrid fimbria protein possesses both the				
CC	immunogenicity and adhesion properties relevant for an efficient live				

CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
 Best Local Similarity 93.0%; Pred. No. 1.2e-121;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGAGCATTGTCGACATGTTGTTCTGCGAGTGTCTGCT 60
DB 1 ATGAACCTTTTAAAGTGGAGCATTGTCGACATGTTGTTCTGCGAGTGTCTGCT 60
QY 61 GGGTCGTTCCACATGGGCGGCGGCGGTAATCATTAACGCGCGCAATAGTCCGCG 120
DB 61 GGGTCGTTCCACATGGGCGGCGGCGGTAATCATTAACGCGCGCAATAGTCCGCG 120
QY 121 CCGGACTCAAGTTCGACATTTATCAGTACGTTCCGCTTAAGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGTTCGACATTTATCAGTACGTTCCGCTTAAGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAAATCTGAAAGCACCATTACCCAGAGCGTTATGTAACGCGCGCAT 240
DB 181 AGCGATGCCCGTAAATCTGAAAGCACCATTACCCAGAGCGTTATGTAACGCGCGCAT 240
QY 241 GTAGGCGGAGGTGGCGGATTAAGTACTATTTGAACGACTCAGATGTTTCAAGAAATAT 300
DB 241 GTAGGCGGAGGTGGCGGATTAAGTACTATTTGAACGACTCAGATGTTTCAAGAAATAT 300
QY 301 GCCACCATGACACGAGTGAACGCTTAAACTCCGATTTACTGTCGCGCAATATGATCAG 360
DB 301 GCCACCATGACACGAGTGAACGCTTAAACTCCGATTTACTGTCGCGCAATATGATCAG 360
QY 361 CTGGTTTACCCGCTGTTTACCATCAATGGCACTGACAGCGTAATGTCGCTCAGGTT 420
DB 361 AATTAACGCGCGCTGTTTATTAAGACCGCATCTGATTCAGCGTAATGTCGCTCAGGTT 420
QY 421 GGTTCGCAACACGCGACGCGCTTAACGAGTATTA 456
DB 421 GGTTCGCAACACGCGACGCGCTTAACGAGTATTA 456

```

RESULT 5

AAC64628 standard; DNA; 456 BP.

```

XX AAC64628;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#7 DNA sequence SEQ ID NO:23.
XX
XX Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Escherichia coli.
XX
XX Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000MO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUYV-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;

```

XX MPI: 2000-672631/65.
 DR P-PSDB: AAB36352.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEB17/TAI) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, GsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 83.9%; Score 382.4; DB 3; Length 456;
 Best Local Similarity 89.9%; Pred. No. 2.4e-114;
 Matches 410; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGAGCATTGTCGACAAATGTTCTGCGAGTGTCTGCT 60
DB 1 ATGAACCTTTTAAAGTGGAGCATTGTCGACAAATGTTCTGCGAGTGTCTGCT 60
QY 61 GGGTCGTTCCACATGGGCGGCGGCGGTAATCATTAACGCGCGCAATATGTTCCGCG 120
DB 61 GGGTCGTTCCACATGGGCGGCGGCGGTAATCATTAACGCGCGCAATATGTTCCGCG 120
QY 121 CCGGACTCAAGTTCGACATTTATCAGTACGTTCCGCTTAAGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGTTCGACATTTATCAGTACGTTCCGCTTAAGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAAATCTGAAAGCACCATTACCCAGAGCGTTATGTAACGCGCGCAT 240
DB 181 AGCGATGCCCGTAAATCTGAAAGCACCATTACCCAGAGCGTTATGTAACGCGCGCAT 240
QY 241 GTAGGCGGAGGTGGCGGATTAAGTACTATTTGAACGACTCAGATGTTTCAAGAAATAT 300
DB 241 GTAGGCGGAGGTGGCGGATTAAGTACTATTTGAACGACTCAGATGTTTCAAGAAATAT 300
QY 301 GCCACCATGACACGAGTGAACGCTTAAACTCCGATTTACTGTCGCGCAATATGATCAG 360
DB 301 GCCACCATGACACGAGTGAACGCTTAAACTCCGATTTACTGTCGCGCAATATGATCAG 360
QY 361 CTGGTTTACCCGCTGTTTACCATCAATGGCACTGACAGCGTAATGTCGCTCAGGTT 420
DB 361 CATGAATGCAACATGCAATATGACCGCATCTGATTCAGCGTAATGTCGCTCAGGTT 420
QY 421 GGTTCGCAACACGCGACGCGCTTAACGAGTATTA 456

```

Db 421 GGTTCGCAACGACGCTAACCATATTA 456

RESULT 6
AAC64622 standard; DNA; 456 BP.

XX ID AAC64622 standard; DNA; 456 BP.
XX
XX AAC64622;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#1 DNA sequence SEQ ID NO:11.
XX
XX Salmoneilla; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmoneilla enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000MO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UTVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36346.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TA) nucleation depended
XX assembly system of strains of Salmoneilla, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmoneilla, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

XX Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 82.5%; Score 376; DB 3; Length 456;
Best Local Similarity 89.0%; Pred. No. 2,9e-112;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGACGATTCGACCAATCGTAGTTTCGCGAGTCTTGCT 60
Db 1 ATGAACCTTTAAAGTGGACGATTCGACCAATCGTAGTTTCGCGAGTCTTGCT 60
QY 61 GCGGTGTTCCACATGAGGCGCGCGGTATATCAACGCGCGGCAATAGTCCGAC 120
Db 61 GCGGTGTTCCACATGAGGCGCGCGGTATATCAACGCGCGGCAATAGTCCGAC 120
QY 121 CCGGACCTCAACGTTGAGCATTATCAAGTCCGTTACGCTGCGCTTCTGCA 180
Db 121 CCGGACCTCAACGTTGAGCATTATCAAGTCCGTTACGCTGCGCTTCTGCA 180
QY 181 AGCGATGCCCTTAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
Db 181 AGCGATGCCCTTAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
QY 241 GTAGGCGAGGTTGCGGATTAATAGTACTATTAAGTCAAGTGGTTTCAGAAATAT 300
Db 241 GTAGGCGAGGTTGCGGATTAATAGTACTATTAAGTCAAGTGGTTTCAGAAATAT 300
QY 301 GCCACCATCGACGAGTGAACGCTTAAAACTCCGATATTACTGCGCAATATGATCAG 360
Db 301 GCCACCATCGACGAGTGAACGCTTAAAACTCCGATATTACTGCGCAATATGATCAG 360
QY 361 CTGGTTACCGCGTGTGTTTACCATGAATGACATGCAATGACCTTAATGTCGTCAGTT 420
Db 361 AATTAACGCGCGGTGTTTATTAATGATCACTGCTTACCGGTGTTTATGATCAATG 420
QY 421 GGTTCGCAACGACGCGCTAACCATGATTTAA 456
Db 421 GCACATCGCAACGCGCGCTAACCATGATTTAA 456

RESULT 7
AAC64626 standard; DNA; 456 BP.

XX ID AAC64626 standard; DNA; 456 BP.
XX
XX AAC64626;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#5 DNA sequence SEQ ID NO:19.
XX
XX Salmoneilla; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmoneilla enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000MO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UTVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36350.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Agfa*, *CsgA* and *Agfa*-homologue fimbriae comprising recombinant
 CC directing recombinant of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant *Agfa*
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant *Agfa* protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match Best Local Similarity 79.3%; Score 361.6; DB 3; Length 456;
 Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTTGCGACATGCTGATGTTCTGCGAGTCTGCT 60
 DB 1 ATGAACTTTTAAAGTGCGAGCATTTGCGACATGCTGATGTTCTGCGAGTCTGCT 60
 QY 61 GGGCGTTCGACAAATGGGCGCGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 DB 61 GGGCGTTCGACAAATGGGCGCGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCGATGCTTCCGTAACGCTGCGTCTGCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCGATGCTTCCGTAACGCTGCGTCTGCTGCA 180
 QY 181 AGCGATGCCGTAATCTGAAACGACCATTTCCAGAGCGGTATGTAACGCGCGCAT 240
 DB 181 AGCGATGCCGTAATCTGAAACGACCATTTCCAGAGCGGTATGTAACGCGCGCAT 240
 QY 241 GTAGGCGAGGTCGCGATTAATGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 300
 DB 241 GTAGGCGAGGTCGCGATTAATGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 300
 QY 301 GCCACCATCGACCGATGGAAGCGTAAACTCCGATTAATCTGCGGCAATATATAT 360
 DB 301 GCCACCATCGACCGATGGAAGCGTAAACTCCGATTAATCTGCGGCAATATATAT 360
 QY 361 CTGGTTCCTGCTGTTTATCCATGAAATGCGACGAGGTAATGTTGCGTCAAGTT 420
 DB 361 CTGGTTCCTGCTGTTTATCCATGAAATGCGACGAGGTAATGTTGCGTCAAGTT 420
 QY 421 GGTTCGCAACGCGGCGGTAATCAAGCGATGCTGATTCAGGTAATGTTGCGTCAAGTT 480
 DB 421 GGTTCGCAACGCGGCGGTAATCAAGCGATGCTGATTCAGGTAATGTTGCGTCAAGTT 480
 QY 481 GGTTCGCAACGCGGCGGTAATCAAGCGATGCTGATTCAGGTAATGTTGCGTCAAGTT 540
 DB 481 GGTTCGCAACGCGGCGGTAATCAAGCGATGCTGATTCAGGTAATGTTGCGTCAAGTT 540

RESULT 8
 AAC64625
 ID AAC64625 standard; DNA; 456 BP.
 AC AAC64625;
 XX
 DT 26-FEB-2001 (first entry)
 XX

DE *Agfa*: PT3#4 DNA sequence SEQ ID NO:17.
 XX *Salmonella*; *Agfa*; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN MO200060102-A2.
 PD 12-OCT-2000.
 XX 05-APR-2000; 2000MO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay W,
 DR WPI; 2000-672631/55.
 PS P-PSDB; AAB36349.
 PT Recombinant *agfa* gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant *Agfa*
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant *agfa* gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Agfa*, *CsgA* and *Agfa*-homologue fimbriae comprising recombinant
 CC directing recombinant of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant *Agfa*
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant *Agfa* protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match Best Local Similarity 77.9%; Score 355.2; DB 3; Length 456;
 Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTTGCGACATGCTGATGTTCTGCGAGTCTGCT 60
 DB 1 ATGAACTTTTAAAGTGCGAGCATTTGCGACATGCTGATGTTCTGCGAGTCTGCT 60
 QY 61 GGGCGTTCGACAAATGGGCGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 DB 61 GGGCGTTCGACAAATGGGCGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCGATGCTTCCGTAACGCTGCGTCTGCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCGATGCTTCCGTAACGCTGCGTCTGCTGCA 180

Db 121 CCGGACTATGATCAAGCTGTTACCGCTGTTTACCCATGAATGGACATGCACTGCA 180
Qy 181 AGCGATGCGCGTAATCTGAACGACCATTAACCGGTTATGTAAGCGCGCCAT 240
Db 181 AGCGATGCGCGTAATCTGAACGACCATTAACCGGTTATGTAAGCGCGCCAT 240
Qy 241 GTAGGCGGAGGTGGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGGCGGAGGTGGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Qy 301 GCCACATGACCAAGTGAAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 360
Db 301 GCCACATGACCAAGTGAAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 360
Qy 361 CTGTTACCGCTGTTGTTACCATGAATGCAATGCAAGCGTAATGTCGTCAGTT 420
Db 361 AATTAACGCGCGCTGTTGTTACCATGAATGCAATGCAAGCGTAATGTCGTCAGTT 420
Qy 421 GGTGTTGGCAACAAAGCGGCGCTTAACCATGTAATTAATTAATTAATTAAT 456
Db 421 GGTGTTGGCAACAAAGCGGCGCTTAACCATGTAATTAATTAATTAATTAAT 456

RESULT 9
AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#8 DNA sequence SEQ ID NO:25.
XX
KM Salmonella; agfa, chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR P-PSDB; AAB36353.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139p; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SFP17/TA) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;
SQ

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 6.1e-105;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1 ATGAACCTTTTAAAGTGGAGCATTTGCGAGCAATCGTAGTTTCTGCGAGTCTGCGT 60
Db 1 ATGAACCTTTTAAAGTGGAGCATTTGCGAGCAATCGTAGTTTCTGCGAGTCTGCGT 60
Qy 61 GCGCGTTCACAAATGGGGGGGGGGGTAATGTAATGTAATGTAATGTAATGTAAT 120
Db 61 GCGCGTTCACAAATGGGGGGGGGGGTAATGTAATGTAATGTAATGTAATGTAAT 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCGCTTCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCGCTTATGATCAG 180
Qy 181 AGCGATCCCGTAATCTGAACGACCATTAACCAAGCGGTTATGTTACGCGCGCAT 240
Db 181 CTGCTTACCGCGTGTGTTTACCATGAATGGCACATCAGGTTATGTTACGCGCGCAT 240
Qy 241 GTAGGCGGAGGTGGGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGGCGGAGGTGGGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Qy 301 GCCACATGACCAAGTGAAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 360
Db 301 GCCACATGACCAAGTGAAGCTTAATTAATTAATTAATTAATTAATTAATTAAT 360
Qy 361 CTGTTACCGCTGTTGTTACCATGAATGCAATGCAAGCGTAATGTCGTCAGTT 420
Db 361 AATTAACGCGCGCTGTTTATCAAGCCGATCTGATTCAGCGTAATGTCGTCAGTT 420
Qy 421 GGTGTTGGCAACAAAGCGGCGCTTAACCATGTAATTAATTAATTAATTAAT 456
Db 421 GGTGTTGGCAACAAAGCGGCGCTTAACCATGTAATTAATTAATTAATTAAT 456

RESULT 10
AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#3 DNA sequence SEQ ID NO:15.
XX
KM Salmonella; agfa, chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
FN WO200060102-A2.
XX

PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 XI WPI; 2000-672631/65.
 XX
 DR P-PSDB; AAB36348.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 CC
 CC Sequence 456 BP, 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;
 XX
 SQ
 Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 2e-104;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
 QY 1 ATGAACTTTTAAAGTGGAGCATTTGGAGCAATCGTACGTTTCTGGAGTCTTGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTTGGAGCAATCGTACGTTTCTGGAGTCTTGCT 60
 QY 61 GGGGTCTTCCACAAATGGGCGGGGGGGAATCATATAAGCGCGGCATTAATGTCGGC 120
 DB 61 GGGGTCTTCCACAAATGGGCGGGGGGGAATCATATAAGCGCGGCATTAATGTCGGC 120
 QY 61 GGGGTCTTCCACAAATGGGCGGGGGGGAATCATATAAGCGCGGCATTAATGTCGGC 120
 DB 61 GGGGTCTTCCACAAATGGGCGGGGGGGAATCATATAAGCGCGGCATTAATGTCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGGCTTGGCTCGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGGCTTGGCTCGCA 180
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGGCTTGGCTCGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCTGGCTTGGCTCGCA 180
 QY 181 AGCGATGCCGTTAAATCTGAAACGACCATTAACCGAGCGGTTATGTTAAACGCGCGAT 240
 DB 181 AGCGATGCCGTTAAATCTGAAACGACCATTAACCGAGCGGTTATGTTAAACGCGCGAT 240
 QY 241 GTAGGCGAGGGTGGGATTAATGACTATGTAAGTGAAGTGGTTTCAAGAAATTAAT 300
 DB 241 GTAGGCGAGGGTGGGATTAATGACTATGTAAGTGAAGTGGTTTCAAGAAATTAAT 300
 QY 301 GCACCATCGACAGTGAACGCTAAATTCGATATTAATGTCGCGCAATATGATCAG 360

DB 301 GCCACCATGACAGTGGAAACGCTAAATCTCGGATTTACTGTCCGCCAATACGCGGT 360
 QY 361 CTGGTACCCGGTGTGTTATCCATGAATGACATGACAGCGTAAATGTCGCTCAGGTT 420
 DB 361 AATAACGCCGCGGTGTTATCAGACCGCATCGATTCAGCGGTAAATGTCGCTCAGGTT 420
 QY 421 GGTTTGGCAACAGCGCACGCGTACCAAGTATTTAA 456
 DB 421 GGTTTGGCAACAGCGCACGCGTACCAAGTATTTAA 456
 RESULT 11
 AAC64630
 ID AAC64630 standard; DNA: 456 BP.
 XX
 AC AAC64630;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#9 DNA sequence SEQ ID NO:27.
 XX
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN W0200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 XI WPI; 2000-672631/65.
 XX
 DR P-PSDB; AAB36354.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SO Sequence 456 BP, 115 A, 116 C, 118 G, 107 T, 0 U, 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 6.8e-104;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGACGATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGACGATTCGAGTTCTGGAGTCTTGCT 60
QY 61 GGGCTCGTTTCAACATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 61 GGGCTCGTTTCAACATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGTTACGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGTTACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATGACTATTTGAACTGATCGATGATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATGACTATTTGAACTGATCGATGATGTTTCAAAATAT 300
QY 301 GCCACATCGACGATGGAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
DB 301 GCCACATCGACGATGGAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
QY 361 CTGCTTACCGTGTGTTTATCCATGAAATGACATGACGATGATGTCGTCAGTT 420
DB 361 AATAACCGCGCGCTGTTTATCCATGAAATGACATGATGTCGTCAGTT 420
QY 421 GGTTCGCAACAGCGGCTAAACGATTAATTA 456
DB 421 GGTTCGCAACAGCGGCTAAACGATTAATTA 456

RESULT 12

AC64627 standard; DNA; 456 BP.
AC AAC64627;
DT 26-FEB-2001 (first entry)
DE Agfa::PT3H6 DNA sequence SEQ ID NO:21.
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
PN MO20060102-A2.
PD 12-OCT-2000.
PF 05-APR-2000; 2000MO-CA000356.
PR 05-APR-1999; 99US-0127888P.
XX (UVI-) UNIV VICTORIA.
PA White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR

DR P-PSDB; AAB36351.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC sequence of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF1/TFP) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SO Sequence 456 BP, 112 A, 113 C, 125 G, 106 T, 0 U, 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 6.8e-104;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGACGATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGACGATTCGAGTTCTGGAGTCTTGCT 60
QY 61 GGGCTCGTTTCAACATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 61 GGGCTCGTTTCAACATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGTTACGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGTTACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATGACTATTTGAACTGATCGATGATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATGACTATTTGAACTGATCGATGATGTTTCAAAATAT 300
QY 301 GCCACATCGACGATGGAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
DB 301 GCCACATCGACGATGGAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
QY 361 CTGCTTACCGTGTGTTTATCCATGAAATGACATGACGATGATGTCGTCAGTT 420
DB 361 AATAACCGCGCGCTGTTTATCCATGAAATGACATGATGTCGTCAGTT 420
QY 421 GGTTCGCAACAGCGGCTAAACGATTAATTA 456
DB 421 GGTTCGCAACAGCGGCTAAACGATTAATTA 456

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 DNA sequence SEQ ID NO:29.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay MW,
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36355.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

SO Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 76.5%; Score 348.8; DB 3; Length 456;
Best Local Similarity 85.3%; Pred. No. 2.3e-103;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

1 ATGAAACTTTTAAAGTGGCAGCATTCGACGACATCGTAGTTCTTGGCAGTGTCTGGCT 60

Db |||||
Qy 1 ATGAAACTTTTAAAGTGGCAGCATTCGACGACATCGTAGTTCTTGGCAGTGTCTGGCT 60
Qy 61 GGGCGTCTTCCAAATGGGGCGGGCGGTAATCATTAACGGCGGGCAATAGTTCGGC 120
Db 61 GGGCGTCTTCCAAATGGGGCGGGCGGTAATCATTAACGGCGGGCAATAGTTCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTTATCAGTACGGTTCGGTCAACGGTCTGCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTTATCAGTACGGTTCGGTCAACGGTCTGCTGCAA 180
Qy 181 AGGATGCCCGGTAAATCTGAAACGACATTACCCAGAGCGGTTATGTAACGGCGCAT 240
Db 181 AGGATGCCCGGTAAATCTGAAACGACATTACCCAGAGCGGTTATGTAACGGCGCAT 240
Qy 241 GTAGGCCAGGTCGGATATATAGTACTATATGAAGTACTCAGAAATGTTTCAAAATAT 300
Db 241 GTAGGCCAGGTCGGATATATAGTACTATATGAAGTACTCAGAAATGTTTCAAAATAT 300
Qy 301 GCCACCATGACCGACGTGAACGCTTAAACCTCCGATATTACTGTGGCCATATATCAG 360
Db 301 GCCACCATGACCGACGTGAACGCTTAAACCTCCGATATTACTGTGGCCATATATCAG 360
Qy 361 CTGGTACCCGCTGTGTATCCCATGAATGGCACATGCAGGCTATGTCGCTCAGGTT 420
Db 361 AATTAACGCCGCGCTGTATATCAGACCGCATCTGATTCAGCGTAAATGTCGCTCAGGTT 420
Qy 421 GGTTTGGCAACAGCCGACGCTAACCAGTATTA 456
Db 421 GGTTTGGCAACAGCCGACGCTAACCAGTATTA 456

RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
XX AAQ73066;
XX
XX 27-AUG-2003 (revised)
XX 25-MAR-2003 (revised)
XX 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
XX
XX Salmonella; agfa; vaccine; genetic immunization; ds.
XX
XX Salmonella enteritidis.
XX
XX Key Location/Qualifiers
XX CDS 1..359
XX /tag= a
XX /note= "Agfa"
XX /tag= d
XX /note= "TAF5 primer (pair with TAF6)"
XX /tag= b
XX /note= "TAF3 primer (pair with TAF4)"
XX /tag= e
XX /note= "TAF6 primer (pair with TAF5)"
XX /tag= c
XX /note= "TAF4 primer (pair with TAF3)"
XX
XX WO9425598-A2.
XX
XX 10-NOV-1994.
XX
XX 26-APR-1994; 94WO-IB000207.
XX
XX 26-APR-1993; 93US-00054452.
XX

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgac.....ccacgcgtacacgtatataa 456

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents NA: *
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/6C.COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	404.8	88.8	456	1	Sequence 58, Appl
2	287.2	63.0	361	1	Sequence 56, Appl
3	34	7.5	1344	3	Sequence 1, Appl
4	34	7.5	1344	4	Sequence 5, Appl
5	33	7.2	7766	4	Sequence 3, Appl
6	31.2	6.8	1491	6	Patent No. 5486473
7	31.2	6.8	1767	1	Sequence 18, Appl
8	31.2	6.8	2284	1	Sequence 21, Appl
9	31	6.8	1750	1	Sequence 1, Appl
10	31	6.8	1750	2	Sequence 15, Appl
11	30.6	6.7	1147	1	Sequence 5, Appl
12	30.6	6.7	1351	1	Sequence 5, Appl
13	30.6	6.7	1351	1	Sequence 5, Appl
14	30.6	6.7	1351	4	Sequence 1303, Ap
15	30.6	6.7	1351	4	Sequence 1552, Ap
16	30.6	6.7	4403765	3	Sequence 2, Appl
17	30.6	6.7	4411529	3	Sequence 9, Appl
18	30.4	6.7	7183	4	Sequence 1, Appl
19	30.4	6.7	7183	4	Sequence 1, Appl
20	30.2	6.6	2007	4	Sequence 153, App
21	30.2	6.6	2007	4	Sequence 153, App
22	30.2	6.6	2007	4	Sequence 153, App
23	30.2	6.6	2007	4	Sequence 153, App
24	30.2	6.6	2007	4	Sequence 153, App
25	30.2	6.6	2007	4	Sequence 153, App
26	30.2	6.6	2148	4	Sequence 154, App
27	30.2	6.6	2148	4	Sequence 154, App

C 28	30.2	6.6	2148	4	US-09-606-421B-154	Sequence 154, App
C 29	30.2	6.6	2148	4	US-09-221-107-154	Sequence 154, App
C 30	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
C 31	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
C 32	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
C 33	29.8	6.5	1612	4	US-09-068-506-2	Sequence 2, Appl
C 34	29.8	6.5	3690	4	US-09-016-434-1234	Sequence 1234, Ap
C 35	29.8	6.5	4035	4	US-09-016-434-1369	Sequence 1369, Ap
C 36	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
C 37	29.4	6.4	423	4	US-09-252-991A-7641	Sequence 7641, Ap
C 38	29.4	6.4	759	2	US-08-853-659A-18	Sequence 22, Appl
C 39	29.4	6.4	918	2	US-08-853-659A-22	Sequence 29, Appl
C 40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
C 41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
C 42	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
C 43	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, App
C 44	29.4	6.4	3660	4	US-09-252-991A-12269	Sequence 12269, A
C 45	29.4	6.4	9208	4	US-09-068-506-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
TITLE OF INVENTION: OF SALMONELLA
CORRESPONDENCE ADDRESSES:
ADDRESSER: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELETYPE: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.5e-11;
Matches 424; Conservative 0; Mismatches 32; Indels 0;

QY 1 ATGAACTTTAAAGTGCAGCATTCGACGAATGTTCTGCGAGTCTGAGCT 60
 DB 1 ATGAACTTTTAAAGTGCAGCATTCGACGAATGTTCTGCGAGTCTGAGCT 60
 QY 61 GCGGTGCTTCCACAAATGGGGCGCGCGGTAATCATTAAGCGCGCAATAGTTCCGGC 120
 DB 61 GCGGTGCTTCCAAATGGGGCGCGCGGTAATCATTAAGCGCGCAATAGTTCCGGC 120
 QY 121 CCGAGTCAACGTTGAGCATTTATCATACGTTCCGTAACGCTGCGCTTCTGCA 180
 DB 121 CCGAGTCAACGTTGAGCATTTATCATACGTTCCGTAACGCTGCGCTTCTGCA 180
 QY 181 AGCGATGCCCGTAAATCTGAAAAGCAACCATTAACCGAGGCGTTAGTAAAGCGCCGCAT 240
 DB 181 AGCGATGCCCGTAAATCTGAAAAGCAACCATTAACCGAGGCGTTAGTAAAGCGCCGCAT 240
 QY 241 GTAGGCGAGGCGGATTAATAGTACTATTAAGTGAAGTCAAGATGTTTCAGAAATAT 300
 DB 241 GTAGGCGAGGCGGATTAATAGTACTATTAAGTGAAGTCAAGATGTTTCAGAAATAT 300
 QY 301 GGCACCATGACACGAGGAAAGCTTAAATCTCCGATTTACTGTGCGCAATATGATCAG 360
 DB 301 GGCACCATGACACGAGGAAAGCTTAAATCTCCGATTTACTGTGCGCAATATGATCAG 360
 QY 361 CTGATTACCCGTGTTGTTTACCATGAAATGGCAGATGCAAGGTAATGTCGTCAGGTT 420
 DB 361 AATAAGCGCGGCTGTTTAAATCAGACCGGATCTGATTCACAGGTAATGTCGTCAGGTT 420
 QY 421 GGTTCGCAACAGCCGCGCTTACCAAGTATTA 456
 DB 421 GGTTCGCAACAGCCGCGCTTACCAAGTATTA 456

RESULT 2

US-08-233-788A-56

Sequence 56, Application US/08233788A
 Patent No. 3633617

GENERAL INFORMATION:
 APPLICANT: Dorian, James L.
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, Karen S.
 APPLICANT: Clouthier, Sharon C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 NUMBER OF INVENTION: 61
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-Apr-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043, 40302
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-4900
 TELEFAX: (206) 682-6031
 TELETYPE: 3723836 SEEDANBERRY
 INFORMATION FOR SEQ ID NO: 56:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 base pairs

TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..357
 US-08-233-788A-56

Query Match 63.0%; Score 287.2; DB 1; Length 361;
 Best Local Similarity 99.0%; Pred. No. 1,8e-90;
 Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 64 GTGCTTCCAAATGGGGCGCGCGGTAATCATTAACGCGGCGCAATAGTTCCGGCCCG 123
 DB 1 GTGCTACCAAGTGGGGCGCGCGGTAATCATTAACGCGGCGCAATAGTTCCGGCCCG 60
 QY 124 GACTCAACGTTGAGCATTTATCATACGTTCCGTAACGCTGCGCTTCTGCAAGC 183
 DB 61 GACTCAACGTTGAGCATTTATCATACGTTCCGTAACGCTGCGCTTCTGCAAGC 120
 QY 184 GATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTAGTAAAGCGCCGATGTA 243
 DB 121 GATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTAGTAAAGCGCCGATGTA 180
 QY 244 GGCAGGCGTGGGATTAATAGTACTATTAAGTGAAGTCAAGATGTTTCAGAAATAT 303
 DB 181 GGCAGGCGTGGGATTAATAGTACTATTAAGTGAAGTCAAGATGTTTCAGAAATAT 240
 QY 304 ACCATGACCACTGGAAGCGCTTAAATCCGATTTACTGTGCGCAATATG 355
 DB 241 ACCATGACCACTGGAAGCGCTTAAATCCGATTTACTGTGCGCAATATG 292

RESULT 3

US-09-120-927-1/c

Sequence 1, Application US/09120927
 Patent No. 6262018

GENERAL INFORMATION:
 APPLICANT: Kim, Jinyun Francis
 APPLICANT: Beer, Steven V.
 TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
 NUMBER OF INVENTION: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 STREET: P.O. Box 1051, Clinton Square
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: PC-DOS/MS-DOS
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/120,927
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/055,108
 FILING DATE: 06-AUG-1977
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldman, Michael L.
 REGISTRATION NUMBER: 30,727
 REFERENCE/DOCKET NUMBER: 19603/1581
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1344 base pairs

RESULT 7

US-08-363-475-18/c

Sequence 18, Application US/08363475
Patent No. 5516679

GENERAL INFORMATION:

APPLICANT: Chiang, Shu-Jen
APPLICANT: Burnett Jr., William V.
TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM
TITLE OF INVENTION: FUSARIUM OXYSPORUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSER: Thomas R. Savitsky
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,475
FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Savitsky, Thomas R.
REGISTRATION NUMBER: 31,661
REFERENCE/DOCKET NUMBER: ON-0134
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252 4956
TELEFAX: (609) 252 4956
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 1767 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1521

US-08-363-475-18

Query Match

Best Local Similarity 6.8%; Score 31.2; DB 1; Length 1767;
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY	278	CTCAGAAATGCTTTGAGAAATATATGCGACATGACGAGTGAACGCTAAAACTCCGATA	337
DB	1391	CCCCCAATGCTTTCTGGCGCAGCCCTTCTCGACCCATTTCAATATCCGAGAACACA	1332
QY	338	TTACTGTGGCGCAATATGATGATGCTGTTACCCGCTGTTGTTACCCATG	385
DB	1331	TTAGTATCTGCCGCTTCTGCGCCCATGTTCCGGCGTTGTTCCATG	1284

RESULT 8

US-08-363-475-21/c

Sequence 21, Application US/08363475
Patent No. 5516679

GENERAL INFORMATION:

APPLICANT: Chiang, Shu-Jen
APPLICANT: Burnett Jr., William V.
TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM
TITLE OF INVENTION: FUSARIUM OXYSPORUM
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:

ADDRESSER: Thomas R. Savitsky

STREET: P.O. Box 4000

CITY: Princeton

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,475
FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Savitsky, Thomas R.
REGISTRATION NUMBER: 31,661
REFERENCE/DOCKET NUMBER: ON-0134
TELECOMMUNICATION INFORMATION:
TELEPHONE: (609) 252 4956
TELEFAX: (609) 252 4956
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 2284 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 241..1836

US-08-363-475-21

Query Match

Best Local Similarity 6.8%; Score 31.2; DB 1; Length 2284;
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY	278	CTCAGAAATGCTTTGAGAAATATATGCGACATGACGAGTGAACGCTAAAACTCCGATA	337
DB	1706	CCCCCAATGCTTTCTGGCGCAGCCCTTCTCGACCCATTTCAATATCCGAGAACACA	1647
QY	338	TTACTGTGGCGCAATATGATGATGCTGTTACCCGCTGTTGTTACCCATG	385
DB	1646	TTAGTATCTGCCGCTTCTGCGCCCATGTTCCGGCGTTGTTCCATG	1599

RESULT 9

US-08-415-823-1

Sequence 1, Application US/08415823
Patent No. 5759538

GENERAL INFORMATION:

APPLICANT: Donovan, William P.
APPLICANT: Tan, Yuding
TITLE OF INVENTION: BACILLUS THURINGIENSIS apt AND npr
TITLE OF INVENTION: PROTEASE GENES AND ALKALINE PROTEASE DEFICIENT AND NEUTRAL
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Panitch Schwarze Jacobs & Nadel, P.C. c/o
STREET: 1601 Market Street, 36th floor
CITY: Philadelphia
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19103-2398COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/415,823

Db 302 CAGACCCAGCATGAAGAGTCTCTGCCATGGCCAGGTTGAGAGTGAATGTTGTGAT 243
QY 274 CTGACTCAGATGTTTTCAGAAATATCCACATCCAGTGGAAAGCTAAACTCC 333
Db 242 GGTCTTCACTCTGGCATAGCGAGATGACATTAATGACAGTGTGGCACAACACC 183
QY 334 GATATTACTGTGGCCCAATATGATCAGCTGTACCCGTTGTTTACCATGAATGGCA 393
Db 182 AATGATGAGACACCAAAATAGATGAATGTGAGAGTGCATTTGCTGTGAGTCAATAGTA 123
QY 394 C 394
Db 122 C 122

RESULT 12

US-07-816-283-5/C
; Sequence 5, Application US/07816283
; Patent No. 5436155
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamada, Yuichiro
; APPLICANT: Saino, Susumu
; TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/816.283
; FILING DATE: 19911231
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniel, C. Steven
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
US-07-816-283-5

Query Match 6.7%; Score 30.6; DB 1; Length 1351;
Best Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 214 CAGAGCGGTATGTTAAGCGGCCGATGTAGGCGAGGTGCGGATTAATGACTAATTGA 273
Db 370 CAGACCCAGCATGAAGAGTCTCTGCCATGGCCAGGTTGAGAGTGAATGTTGTGAT 311
QY 274 CTGACTCAGATGTTTTCAGAAATATCCACATCCAGTGGAAAGCTAAACTCC 333
Db 310 GGTCTTCACTCTGGCATAGCGAGATGACATTAATGACAGTGTGGCACAACACC 251
QY 334 GATATTACTGTGGCCCAATATGATCAGCTGTACCCGTTGTTTACCATGAATGGCA 393
Db 250 AATGATGAGACACCAAAATAGATGAATGTGAGAGTGCATTTGCTGTGAGTCAATAGTA 191
QY 394 C 394

Db 190 C 190

RESULT 13

US-08-417-103-5/C
; Sequence 5, Application US/08417103
; Patent No. 5723299
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamada, Yuichiro
; APPLICANT: Saino, Susumu
; TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417.103
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/816.283
; FILING DATE: 01-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 83..1189
US-08-417-103-5

Query Match 6.7%; Score 30.6; DB 1; Length 1351;
Best Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 214 CAGAGCGGTATGTTAAGCGGCCGATGTAGGCGAGGTGCGGATTAATGACTAATTGA 273
Db 370 CAGACCCAGCATGAAGAGTCTCTGCCATGGCCAGGTTGAGAGTGAATGTTGTGAT 311
QY 274 CTGACTCAGATGTTTTCAGAAATATCCACATCCAGTGGAAAGCTAAACTCC 333
Db 310 GGTCTTCACTCTGGCATAGCGAGATGACATTAATGACAGTGTGGCACAACACC 251
QY 334 GATATTACTGTGGCCCAATATGATCAGCTGTACCCGTTGTTTACCATGAATGGCA 393
Db 250 AATGATGAGACACCAAAATAGATGAATGTGAGAGTGCATTTGCTGTGAGTCAATAGTA 191
QY 394 C 394
Db 190 C 190

RESULT 14

US-09-016-434-1303/c
; Sequence 1303, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Selhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HEREMITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1303:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g307435
; US-09-016-434-1303
Query Match 6.7%; Score 30.6; DB 4; Length 1351;
Best Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;
QY 214 CAGAGCGGTTATGTAACGGCGCGATGAGCCAGGTCGGAATATAGTACTATTGAA 273
Db |||||
QY 370 CAGAGCCAGCATGTAAGAGCTCATCTGCATGCGACGATGTAAGTAAATGTTGGTAT 311
Db |||||
QY 274 CTGACTCAGATGTTTCAGAAATTAATGCCATCGACGATGGAAGCTTAAACTCC 333
Db |||||
QY 310 GGTCTTCTCTTGCGATAGCGGAGATGACATTAATGCAAGTGTGTCACACAAACC 251
Db |||||
QY 334 GATATTACTGCGCCATATGATGACGCTGTTACCCGTGTTTACCAGTAATGGA 393
Db |||||
QY 250 AATGATGACAGCAACAATAATGATGAATGAGAGCTGCAATGCTTGTCAAGTATGTA 191
Db |||||
QY 394 C 394
Db 190 C 190

RESULT 15
US-09-134-000C-1552/c
; Sequence 1552, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1552
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; US-09-134-000C-1552
Query Match 6.7%; Score 30.6; DB 4; Length 1797;
Best Local Similarity 65.2%; Pred. No. 2.3;
Matches 45; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
QY 361 CTGATTACCGGTGTTGTTTACCATGAATGACATGCAAGCTTAATGTCGTCAGGTT 420
Db |||||
QY 721 CTGCTTTTCTTCTGCTTTCATGAATTCACATATAAGCGCAATTAATCATTAATTG 662
Db |||||
QY 421 GTTTGGC 429
Db 661 GTTGTGGC 653

Search completed: March 16, 2004, 04:37:25
Job time : 53.6647 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds

(without alignments) 4179.927 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgc.....ccacgcgttaaccagttatataa 456

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues 4865114

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34.4	7.5	2751	15 US-10-398-221-3563	Sequence 3563, Ap
2	34.2	7.5	616	12 US-10-424-599-88945	Sequence 88945, A
3	34	7.5	1344	9 US-09-835-684-6	Sequence 6, Appli
4	34	7.5	1344	9 US-09-880-371-6	Sequence 6, Appli
5	34	7.5	1344	9 US-09-879-248-5	Sequence 5, Appli
6	34	7.5	1344	14 US-10-010-390-6	Sequence 5, Appli
7	34	7.5	1344	15 US-10-441-736-5	Sequence 5, Appli
8	34	7.5	1625	15 US-10-369-493-26980	Sequence 26980, A
9	33.2	7.3	1662	14 US-10-156-761-3463	Sequence 3463, Ap
10	33.2	7.3	9025608	14 US-10-156-761-1	Sequence 1, Appli
11	33	7.2	7766	12 US-10-222-566-3	Sequence 3, Appli
12	33	7.2	7766	14 US-10-222-162-3	Sequence 3, Appli
13	33	7.2	7766	14 US-10-143-024-3	Sequence 3, Appli
14	33	7.2	368004	9 US-09-949-654-3	Sequence 3, Appli
15	32.8	7.2	1776	12 US-10-282-122A-24385	Sequence 24385, A

C 16	32.6	7.1	596	14 US-10-184-644-310	Sequence 310, App
C 17	32.6	7.1	596	14 US-10-184-634-310	Sequence 310, App
C 18	32.6	7.1	596	14 US-10-063-685-100	Sequence 100, App
C 19	32.2	7.1	568	14 US-10-333-631-1	Sequence 1, Appli
C 20	32.2	7.1	708	12 US-10-282-122A-27953	Sequence 27953, A
C 21	32.2	7.1	972	14 US-10-333-631-4	Sequence 4, Appli
C 22	32	7.0	983	9 US-09-764-864-465	Sequence 465, App
C 23	31.8	7.0	750	14 US-10-184-644-104	Sequence 104, App
C 24	31.8	7.0	750	14 US-10-184-634-104	Sequence 104, App
C 25	31.8	7.0	978	15 US-10-369-493-24285	Sequence 24285, A
C 26	31.8	7.0	4589	10 US-09-764-891-8845	Sequence 8845, Ap
C 27	31.8	7.0	4590	10 US-09-764-891-8846	Sequence 8846, Ap
C 28	31.8	7.0	24081	14 US-10-132-134-13	Sequence 13, Appli
C 29	31.8	7.0	52101	14 US-10-132-134-1	Sequence 1, Appli
C 30	31.6	6.9	4830	12 US-10-282-122A-27869	Sequence 27869, A
C 31	31.6	6.9	6003	13 US-10-010-501-9	Sequence 9, Appli
C 32	31.4	6.9	269	9 US-09-923-876-5903	Sequence 5903, Ap
C 33	31.4	6.9	269	11 US-09-923-876-5903	Sequence 5903, Ap
C 34	31.4	6.9	1287	15 US-10-369-493-46980	Sequence 46980, A
C 35	31.4	6.9	2007	12 US-10-425-114-31950	Sequence 31950, A
C 36	31.4	6.9	2074	12 US-10-425-114-27886	Sequence 27886, A
C 37	31	6.8	2886	9 US-09-801-368-131	Sequence 131, App
C 38	30.8	6.8	421	10 US-09-918-995-33997	Sequence 33997, A
C 39	30.6	6.7	1110	11 US-09-826-509-570	Sequence 570, App
C 40	30.6	6.7	1110	14 US-10-225-567A-313	Sequence 313, App
C 41	30.6	6.7	1164	12 US-10-282-122A-13500	Sequence 13500, A
C 42	30.6	6.7	1351	15 US-10-305-720-1303	Sequence 1303, Ap
C 43	30.6	6.7	1510	15 US-10-292-798-677	Sequence 677, App
C 44	30.6	6.7	1608	9 US-09-815-242-3923	Sequence 3923, Ap
C 45	30.6	6.7			

ALIGNMENTS

RESULT 1

US-10-398-221-3563

Sequence 3563, Application US/10398221

Publication No. US20040018514A1

GENERAL INFORMATION:

APPLICANT: KUNST, Frederik

TITLE OF INVENTION: Listeria innocua, genome and applications

FILE REFERENCE: 344 702 - US

CURRENT FILING DATE: 2003-03-27

PRIOR APPLICATION NUMBER: PCT/FR 01/03 061

PRIOR FILING DATE: 2001-10-04

PRIOR APPLICATION NUMBER: FR 00/12 697

PRIOR FILING DATE: 2000-10-04

NUMBER OF SEQ ID NOS: 4025

SOFTWARE: PatentIn version 3.0

SEQ ID NO 3563

LENGTH: 2751

TYPE: DNA

ORGANISM: Listeria monocytogenes 4b

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)..(end)

OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u

US-10-398-221-3563

Query Match 7.5%; Score 34.4; DB 15; Length 2751;

Best Local Similarity 52.0%; Pred. No. 1.1;

Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Query

238 GATGATGACGAGGTGGGATTAATGACTTATGAACTGACGATGATGTTTCAAAAT 297

DB

321 GATGATGATGAAGAGGAAATATTTACTATGACCAACCGTTTCTGATATACATAT 380

Query

298 AATGCCACATCGACGATGAAAGCTTAAATCTCCATATTAATGATGATGATATGAT 357

DB

381 AAGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 440

OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

Query Match	Score	DB	Length
Best Local Similarity	7.5%	34.2	12
Matches	50.9%		
81; Conservat...	Pred. No. 0.56		616

[illegible]

```

RESULT 3
US-09-835-684-6/c
; Sequence 6, Application US/09835684
; Patent No. US20020019337A1
; GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Qiu, Dewen
APPLICANT: Remick, Dean
TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
FILE REFERENCE: 21829/71
CURRENT APPLICATION NUMBER: US/09/835,684
CURRENT FILING DATE: 2001-04-16
PRIOR APPLICATION NUMBER: 60/198,359
NUMBER OF SEQ ID NOS: 12
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 1344
TYPE: DNA
ORGANISM: Etmunia amylovora
US-09-835-684-6

```

Query Match	7.5%	Score 34;	DB 9;	Length 1344;
Best Local Similarity	54.9%;	Pred. No. 1;		
Matches 67;				

QY	52	GCTCTGCTGGCGATGTTCCACATATGGGCGCGCGCGATATCATATAGGCGCGCGCAT	112
Db	209	GCCTCCGGTTCGGCATTTACTGATTTGGGAGATAGAGATTTAAAGTTCCGCAT	111
QY	112	AGTCCGCGCGCGGACTCAAGCTTATGACATTTATCACTACGTTCCGCTAAGCTCGCGTT	150
Db	149	AATTAGGCATTTCCTCATATGTTTCCCATGATGGGTTGTGCCCAAGCAGAAATTT	171
QY	172	GC 173	
Db	89	GC 88	

RESULT 4
US-09-880-371-6/c

Sequence 6, Application US/098803771
Patent No. US20020059658A1
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Derocher, Jay
TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
FILE REFERENCE: 2129/91
CURRENT APPLICATION NUMBER: US/09/880,371
CURRENT FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: 60/211,585
PRIORITY FILING DATE: 2000-06-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1344
TYPE: DNA
ORGANISM: *Erynia amylovora*
IS-09-880-371-6

```
Query Match      7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%;
Matches 67; Pred. No 1.
```

[illegible]

RESULT 5
US-09-879-248-5/c

Sequence 5, APPLICATION US/09879248
Patent No. US2002062500A1
GENERAL INFORMATION:
APPLICANT: Fan, Hao
APPLICANT: Mei, Zhong-Min
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
FILE REFERENCE: 21829/81
CURRENT APPLICATION NUMBER: US/09/879,248
PRIOR APPLICATION NUMBER: 2001-06-12
PRIOR FILING DATE: 2000-06-16
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match
Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGCGCTGCTTCACATGGGCGCGCGGTAATCATTAACGGCGCGCAAT 111
DB 209 GCTCCGGTGGCGGATTAATGATGATGCGGATGACGATTAACAGTTCCGCCAAT 150
QY 112 AGTTCGGCGCGGAGCTACAGCTTATGAGCATTTATGAGTTCGCTTACGCTGCGCTT 171
DB 149 AATTGAGCATTTGCTCAATGATGTTGCGGATGATGAGTGTGCGCCACAGCAAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 6

US-10-010-390-6/c
Sequence 6, Application US/10010390
Publication No. US20030104979A1
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Leon, Ernesto
APPLICANT: Oviedo, Agustín
TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
FILE REFERENCE: 21829/111
CURRENT APPLICATION NUMBER: US/10/010.390
CURRENT FILING DATE: 2001-11-05
PRIOR APPLICATION NUMBER: 60/248,169
PRIOR FILING DATE: 2000-11-13
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 6
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-10-010-390-6

Query Match
Best Local Similarity 7.5%; Score 34; DB 14; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGGCTGCGCTGCTTCACATGGGCGCGCGGTAATCATTAACGGCGCGCAAT 111
DB 209 GCTCCGGTGGCGGATTAATGATGATGCGGATGACGATTAACAGTTCCGCCAAT 150
QY 112 AGTTCGGCGCGGAGCTACAGCTTATGAGCATTTATGAGTTCGCTTACGCTGCGCTT 171
DB 149 AATTGAGCATTTGCTCAATGATGTTGCGGATGATGAGTGTGCGCCACAGCAAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 7

US-10-441-736-5/c
Sequence 5, Application US/10441736
Publication No. US20040016029A1
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Schading, Richard L.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
TITLE OF INVENTION: RESISTANCE
FILE REFERENCE: 21829/203 (EBC-003)

CURRENT APPLICATION NUMBER: US/10/441.736
CURRENT FILING DATE: 2003-05-20
PRIOR APPLICATION NUMBER: 60/107,243
PRIOR FILING DATE: 1998-11-05
PRIOR APPLICATION NUMBER: 09/431,614
PRIOR FILING DATE: 1999-11-02
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 5
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-10-441-736-5

Query Match
Best Local Similarity 7.5%; Score 34; DB 15; Length 1344;
Best Local Similarity 54.9%; Pred. No. 1;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGCGCTGCTTCACATGGGCGCGCGGTAATCATTAACGGCGCGCAAT 111
DB 209 GCTCCGGTGGCGGATTAATGATGATGCGGATGACGATTAACAGTTCCGCCAAT 150
QY 112 AGTTCGGCGCGGAGCTACAGCTTATGAGCATTTATGAGTTCGCTTACGCTGCGCTT 171
DB 149 AATTGAGCATTTGCTCAATGATGTTGCGGATGATGAGTGTGCGCCACAGCAAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 8

US-10-369-493-26980/c
Sequence 26980, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369.493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 26980
LENGTH: 1625
TYPE: DNA
ORGANISM: Neurospora crassa
US-10-369-493-26980

Query Match
Best Local Similarity 7.5%; Score 34; DB 15; Length 1625;
Best Local Similarity 61.1%; Pred. No. 1.1;
Matches 55; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 290 TCAGAAATTAATGACCACTGACAGTGAAGCAAGCTAAATCTCGATATTACTGCGCC 349
DB 395 TCGGAATTTCTGTCAAGCTGCGCCACAGCAATGTCGAACTTGATTAACGTCGCGC 336
QY 350 AATATGATCAGCTGCTTACCGGTGTTGTTA 379
DB 335 AATCTGATGGGGATGATGCCCTTGTCAATA 306

RESULT 9

US-10-156-761-3483/c
Sequence 3483, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:

Query Match	7.3%	Score 33.2;	DB 14;	Length 1662;
Best Local Similarity	52.1%	Pred. No. 2.1;		
Matches	74;	Conservative	0;	Mismatches 2

RESULT 10
US-10-156-761-1
; Sequence 1, Application US/10156761;
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICATION:

APPLICANT: UMURA, SATOSHI
 APPLICANT: IKEDA, HARUO
 APPLICANT: ISHIKAWA, JUN
 APPLICANT: HORIKAWA, HIROSHI
 APPLICANT: SHIBA, TADAYOSHI
 APPLICANT: SAKAKI, YOSHITAKA
 APPLICANT: HATTORI, MASAHIRA
 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-262
 CURRENT APPLICATION NUMBER: US/10/156,761
 CURRENT FILING DATE: 2002-05-29
 PRIOR APPLICATION NUMBER: JP 2001-204089
 PRIOR FILING DATE: 2001-05-30
 PRIOR APPLICATION NUMBER: JP 2001-272697
 PRIOR FILING DATE: 2001-08-02
 NUMBER OF SEQ ID NOS: 15109
 SEQ ID NO 1
 LENGTH: 9025608
 TYPE: DNA
 ORGANISM: Streptomyces avermitilis
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (4167715)
 OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

[illegible]

```

RESULT 11
US-10-222-566-3
; Sequence 3, Application US/10222566
; Publication No. US20040044192A1
GENERAL INFORMATION:
APPLICANT: NORRIS, STEVEN J.
APPLICANT: JING-REN, ZHANG
APPLICANT: HARDHAM, JOHN M.
APPLICANT: HOWELL, JERRILYN K.
APPLICANT: BARBOUR, ALAN G.
APPLICANT: WEINSTOCK, GEORGE M.
TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
FILE REFERENCE: UTSJ:234USD3
CURRENT APPLICATION NUMBER: US/10/222,566
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: 09/125,619
PRIOR FILING DATE: 1992-01-27
NUMBER OF SEQ ID NOS: 50
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3
LENGTH: 7766
TYPE: DNA
ORGANISM: Borrelia burgdorferi
FEATURE:
NAME/KEY: modified_base
LOCATION: (127)
OTHER INFORMATION: R = A OR G
US-10-222-566-3

```

	Query Match Best Local Similarity	7.2%; 47.8%;	Score 33; Pred. No. 5.8;	DB 12;	Length 7766;
	Matches	96;	Conservative	0;	Mismatches 105; Indels 0; Gaps 0;
QY	157	GCTAACGCTGCGGCTGCTCTGTCGCAAGCGATCCCGCTAACTGAAACGACCATTCACCG			216
Db	4432	GTTAACGCGCTGCGGCTGCTGTGTGACGAGATGTGAGAGAGCTTCACAGGCTCAAAATCCG			4492
QY	217	AGCGTATTATGTGTAAGCGCGCCGATGTATAGCCAGGCTGCGGATATATGATCTATTAAGTC			276
Db	4492	ATTGCTGCTGCTATTGTGGAGAGGTAATGAGAGTGGCGGATTTTGTATAGATGATGATG			4551
QY	277	ACTCAGATGTTTCAGAAATATATGACCACCATGACCACTGGAAGCGCTTAAAACTCCGAT			336
Db	4552	AAGAAAGATGATCAGATTGTGCTGCTGCTATTGCTTTGAGGGGATGCTTAAGATGGAAG			4611
QY	337	ATTACTGTGCGCAATATGAT			357
Db	4612	TTTGCTGTGAAGATATATGAT			4632

RESULT 12
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.


```

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert

```

```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 24385
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Listeria monocytogenes
US-10-282-122A-24385

```

```

Query Match          7.2%; Score 32.8; DB 12; Length 1776;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 238 GATGAGCCCGGGTGGCGATATAGTACTATTGACTGACTGAGATGGTTTCAGAAAT 237
DB 775 GATGTTGTCAAGAAAGCAAAAATATACCATTTGAAAAACGGTTCTCGTATGACATAT 834
QY 298 AATGCCACCATGCGACGAGTGAAGCTAATAAATCCGATATTACTGTGCGCCATATGAT 357
DB 835 AAAAGAGCAATGATCGTTTGGGCAATGATTAACCGGATATTCGTTTGGTTAGAAATTA 894
QY 358 CAGCTGTTAACCGCGTGTGTATCCCATG 385
DB 895 CAAATGTATCCGATGCTGTTAAAGATG 922

```

Search completed: March 17, 2004, 08:15:58
Job time : 415.736 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgc.....ccacgctcaccagctatcaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: em_estba:*
10: em_estba:*
11: em_estba:*
12: em_estba:*
13: em_estba:*
14: em_estba:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*
28: em_estba:*
29: em_estba:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	226.2	49.6	680	12	BU618688
C 2	36	7.9	523	12	BU334624
C 3	35.6	7.8	744	14	CF446143
C 4	34.8	7.6	549	28	B2295150

Result No.	Score	Query Match	Length	ID	Description
C 5	34.8	7.6	594	12	BU335653
C 6	34.8	7.6	813	28	B2204853
C 7	34.6	7.6	540	14	CB210843
C 8	34.6	7.6	556	9	AV976624
C 9	34.4	7.5	539	13	BQ141266
C 10	34.4	7.5	560	10	AW686303
C 11	34.4	7.5	627	10	AW573882
C 12	34.2	7.5	456	12	BU331339
C 13	34.2	7.5	524	12	BU367372
C 14	34.2	7.5	547	12	BU333958
C 15	34.2	7.5	553	12	BU336903
C 16	34.2	7.5	569	12	BU334047
C 17	34.2	7.5	574	12	BU330328
C 18	34.2	7.5	607	12	BU355766
C 19	34.2	7.5	625	12	BU369922
C 20	34.2	7.5	629	12	BU335378
C 21	34.2	7.5	635	12	BU397391
C 22	34.2	7.5	645	12	BU392753
C 23	34.2	7.5	670	12	BU336637
C 24	34.2	7.5	1036	12	BG671547
C 25	34	7.5	234	10	BB581987
C 26	34	7.5	500	14	CA711477
C 27	34	7.5	512	10	BB757100
C 28	33.8	7.4	298	14	F19524
C 29	33.8	7.4	365	9	AA746477
C 30	33.8	7.4	558	14	CD374421
C 31	33.8	7.4	559	13	BU765156
C 32	33.8	7.4	941	29	CG178280
C 33	33.8	7.4	1098	29	CNS02GQP
C 34	33.6	7.4	712	14	CF437982
C 35	33.6	7.4	811	28	BZ469000
C 36	33.4	7.3	402	13	BU767674
C 37	33.4	7.3	505	6	AL809250
C 38	33.4	7.3	536	13	BU767674
C 39	33.4	7.3	546	13	BU767674
C 40	33.4	7.3	699	12	BI959328
C 41	33.4	7.3	715	13	BU444683
C 42	33.2	7.3	589	28	BZ300068
C 43	33.2	7.3	629	9	AU295325
C 44	33.2	7.3	788	9	AU139321
C 45	33.2	7.3	1646	10	BF579120

ALIGNMENTS

RESULT 1
LOCUS BU618688/c
DEFINITION BU618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION BU618688
VERSION BU618688.1 GI:37256713
KEYWORDS EST.

SOURCE
ORGANISM Xenopus laevis (African clawed frog)

REFERENCE
AUTHORS Kohara,Y., Terasaka,C., Mochii,M., Ueno,N., Shin-i.T. and
Kohara,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
CONTACT Tadasu Shin-i
CENTER For Genetic Resource Information
NATIONAL INSTITUTE OF GENETICS
111 Yata, Mishima, Shizuoka 411-8540, Japan
TEL: 81-559-81-6856
FAX: 81-559-81-6855
EMAIL: tehin@genie.nig.ac.jp
URL: The information of this clone is available through the following

BU335653 BU335653
B2204853 CH230-390
CB210843 OM101123
AV976624 AV976624
BQ141266 NF017G03P
AW686303 NF036D06N
AW573882 EST316473
BU331339 BU331339
BU367372 BU367372
BU333958 BU333958
BU336903 BU336903
BU334047 BU334047
BU330328 BU330328
BU365766 BU365766
BU369922 BU369922
BU335378 BU335378
BU397391 BU397391
BU392753 BU392753
BU336637 BU336637
BG671547 DRNBUB05
BB581987 BB581987
CA711477 wdk2c.pko
BB757100 BB757100
F19524 HSPD04128 H
AA746477 nw62e02.s
CD374421 TMBMFC2H
BU765156 aa823f11.
CG178280 PU0DF79TD
AL196666 Telradon
CF437982 EST674327
BZ469000 BOOAO8TF
BU767674 SUBBR08
AL809250 Ttlicum
BU767674 SUBBR08
BU767674 SUBBR08
BI959328 HVSME001
BU444683 BU3765715
BZ300068 KDS010.PI
AU295325 AU295325
AU139321 AU139321
BF579120 602096054

FEATURES
source location/Qualifiers
1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL18b22"
/issue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN
http://xenopus.nibb.ac.jp.

Query Match 49.6%; Score 226.2; DB 12; Length 680;
Best Local Similarity 68.4%; Pred No. 3.5e-53;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGCGACATTCGACCAATCGTATTTCTGCGAGTGGCTGGCT 60
DB |||||||
593 ATGAACCTTTTAAAGTGCGACATTCGACCAATCGTATTTCTGCGAGTGGCTGGCT 60
QY 61 GCGCTGCTCCACATGCGCGCGCGGTATCATTAACGCGCGCAATAGTTCGGC 120
DB |||||||
533 GGTGTTGTTCCAGTACGCGCGCGGTATCATTAACGCGCGCAATAGTTCGGC 120
QY 121 CCGGACTCAAGCTTATTCAGTACGCGGTTCGCTAACGCTGCGTCTGCGCA 180
DB |||||||
473 CCAATTTGAGCTGCAATTTACAGTACGCGGTTCGCTAACGCTGCGTCTGCGCA 180
QY 181 AGCGATGCGCTTAATCGAAAGCACTTACCCAGAGCGGTATGTAACGCGCGCAT 240
DB |||||||
413 ACTGATGCGCTTACTGCTGCTTATTCACCAAGCGCGGTATGTAACGCGCGCAT 240
QY 241 GTAGCGCGGTGCGGTATTCAGTACGCTTACTGCTGCTTATTCAGTACGCTTACT 300
DB |||||||
353 GTAGCGCGGTGCGGTATTCAGTACGCTTACTGCTGCTTATTCAGTACGCTTACT 300
QY 301 GCCACCATCGACCGTAAAGCTTAAAGCTCCGATTTACTGTCGCGCAATATGATCAG 360
DB |||||||
293 GCTACTCTGATAGTGAAGCGCAAAATTTCTGAATGACGCTTGGTGGT 234
QY 361 CTGGTACCGCTGTTGTTTACCATGAATGCGACATGCGTAAAGTGGTGGT 234
DB |||||||
233 GCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 421 GGTGTTGCGACACCGCGCTAACCGATTTAA 456
DB |||||||
173 GGTGTTGCGACACCGCGCTAACCGATTTAA 138

RESULT 2
BU334624 523 bp mRNA linear EST 05-MAR-2002
LOCUS BU334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
DEFINITION BU334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
ACCESSION BU334624
VERSION BU334624.1 GI:19164754
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehini@genes.nig.ac.jp.

FEATURES
source location/Qualifiers
1. .523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47006"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"

ORIGIN

Query Match 7.9%; Score 36; DB 12; Length 523;
Best Local Similarity 55.5%; Pred. No. 30;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 238 GATGTAGCGCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
DB |||||||
398 GATGTAGCGCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 297
QY 298 AATGCCACATCGACCGTGAACCGTAAACCTCCGATTTACTGTCGCGCAATATGA 356
DB |||||||
458 AATGCCACATCGACCGTGAACCGTAAACCTCCGATTTACTGTCGCGCAATATGA 356

RESULT 3
CF446143 744 bp mRNA linear EST 04-SEP-2003
LOCUS CF446143 EST682488 normalized cDNA library of onion Allium cepa cDNA clone
DEFINITION AC446143
ACCESSION CF446143
VERSION CF446143.1 GI:34468845
KEYWORDS EST.
SOURCE Allium cepa (onion)
ORGANISM Allium cepa (onion)
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
Allium.
AUTHORS Haver, M.J., Cheng, F., Van Aken, S., Uterback, T. and Town, C.D.
TITLE Expressed Sequence Tags from a normalized library of mixed onion tissues (Allium cepa)
JOURNAL Unpublished (2003)
COMMENT Contact: Haver M
Department of Horticulture
USDA-ARS and University of Wisconsin
1575 Linden Drive, Madison, WI 53706, USA
Tel: 608-262-1830
Fax: 608-262-4743
Email: mjhaver@facstaff.wisc.edu
TIGR sequence name AC446143.1
http://haverlab.hort.wisc.edu
Seq primer: CAG GAA ACA GCT ATG ACC.
Location/Qualifiers
1. .744
/organism="Allium cepa"
/mol_type="mRNA"
/cultivar="Red Creole (bulbs), unknown (callus), Espano &
Texas Legend (roots)"
/db_xref="taxon:4679"
/clone="AC446143"
/issue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/note="Vector: pCMVSPORT6.1-ccdb (Invitrogen); Site 1:
from callus, roots, and young bulbs were combined to
synthesize the library. Normalization to enrich for
low-copy transcripts was performed by proprietary
techniques of Invitrogen."

ORIGIN

Query Match 7.8%; Score 35.6; DB 14; Length 744;
Best Local Similarity 47.7%; Pred. No. 46;

	Matches	104;	Conservative	0;	Mismatches	114;	Indels	0;	Gaps	0
Qy	224	CGCCGATGAGCC	CGGGTCCGGATTAATATGTA	CTTTGGA	CTGACTCAGATG	TTTTCAG	293			
Db	382	CGAAGCTA	AAAGCA	CAAACTTTA	TAGCAATTTATATCAT	CTGACTGCTATGATATCG	323			
Qy	294	AAATATATGCAC	CATCGAC	CAGTGGAA	CGCTAA	AAACTCCGATTA	CTGTCGGCCAA	274		
Db	322	AAATTAACA	TAGTA	TACAGCA	MAATGATCTCTTAA	AGGTGTGTTAGTACTATCA	TAAATGA	263		
Qy	354	TGATCAGCTG	TTTACCCCTG	TTGTTTACC	CATGA	AAATGCACTGCAAGCGTA	TGTCG	413		
Db	262	AGAGCGAG	AGAAACCA	CTCTTTTATG	ATGATCAATCTGTTCA	TATATGTTGC	AGAGAGG	203		
Qy	414	TCAGGTG	TTTTGGCA	CAACGCCA	CGGCTAAC	AGT	451			
Db	202	AAGATT	TAA	CACTTCAAC	CAAAATGCGCTCG	AAGT	165			

RESULT 4	B2295150/c	549 bp	DNA	linear	GSS 31-OCT-2007
LOCUS	B2295150/c				
DEFINITION	B2295150				
ACCESSION	CG1320.r1				Candida glabrata Random Genomic Library
VERSION	B2295150				Candida glabrata
KEYWORDS	B2295150.1				genomic clone CG1320, genomic survey sequence.
SOURCE	GSS:				
ORGANISM	Candida glabrata				
	Candida glabrata				
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
	Saccharomycetales; mitosporic Saccharomycetales; Candida.				
REFERENCE	1 (bases 1 to 549)				
AUTHORS	Wong,S., Pares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.				
TITLE	Evidence from comparative genomics for a complete sexual cycle in				
	the asexual pathogenic yeast Candida glabrata				
JOURNAL	Genome Biol.				
MEDLINE	22508158				
PUBMED	12620120				
COMMENT	Contact: Wong S				

Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmaid ends.
Location/Qualifiers
1. .549

```

/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS_138"
/db_xref="taxon:5478"
/clone="CG1320"
/clone_11b="Candida glabrata Random Genomic Library"
ORIGIN

```

Query Match	7.6%	Score 34.8	DB 28	Length 549
Similarity	52.0%	Pred. No. 68		
Best Local				
Matches	78	Conservative	0	Mismatches 72; Indels 0; Gaps 0;
QY	222	TTATGCTAACGGCGCCGATGTAGGCGCAGGTATATAGTATTAAGTACATCA	281	
Db	291	TTATAGTTTTTGCTCTCAAGGCGCAATTTTCAGTAAAAAGCTGAAATTTTTCGA	232	
QY	282	GAATGCTTCAGAAATAATGCCACCATGACACAGTGAACGCTAAAACTCGATATTAC	341	
Db	231	GAATATTACAGTGTGTATACGAAACACATCTAGATCGAACCACTCTATATCAG	172	
QY	342	TGTGGCCCAATATGATCAGCTGTATCCG	371	
Db	171	TGCTGCCAAGAAATATCCTTTTGATTAACG	142	

RESULT 5									
LOCUS	BJ335653								
DEFINITION	BJ335653	594 bp	mRNA	linear	EST 05-MAR-2002				
ACCESSION	BJ335653	Dictyostelium discoideum cDNA library, AF							
VERSION	BJ335653	dictoideum cDNA clone dda51h06 5', mRNA sequence.							
KEYWORDS	BJ335653..1	GI:19165783							
SOURCE	EST.								
ORGANISM	Dictyostelium discoideum								
REFERENCE	Dictyostelium discoideum								
AUTHORS	Eukaryota; Mycelozoa; Dictyostellida; Dictyostelium.								
TITLE	1 (bases 1 to 594)								
JOURNAL	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.								
COMMENT	Full length cDNA of Dictyostelium discoideum at the aggregation stage								
	Contact : Tadasu Shin-i								

Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: rehin@genes.nig.ac.jp.
Location/Qualifiers
1. .594

```

/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda51h06"
/sex="mat A"
/dev_stage="aggregation stage"
/clone_1b="Dictyostelium discoideum cDNA library, Af"

```

	Query Match	Best Local Similarity	Score 34.8%	DB 12	Length 594
Matches	66	Conservative	55.5%	0	Indels 0
			Pred. No. 70		Gaps 0
			Mismatches	53	
QY	238	GATGTAGCCGCGGTGGCGATTAATAGACTATTGGAATGCTGACCTGAGATGGTTTCAGAAAT	297		
DB	223	GATGAAGATGATGATTTGCGAGATAGATATATTGGATTCATAGAAATGAGACAAATATAT	282		
QY	298	AATGCCACCATCGACCGCTGGAACGCTAAAACTCGCATTTACTGTGGCCCAATATGCA	356		
DB	283	AATTAACAACAACAACAACAATTAATAAATAAATAATATTTATATAAAGATGATGAATGCA	341		

RESULT 6	
LOCUS	BZ204853
DEFINITION	BZ204853 813 bp DNA linear GSS 11-OCT-2002
CH30-390D21, TV CHORI-230 Segment 2 Rattus norvegicus genomic clone.	
CH230-390D21, genomic survey sequence.	
ACCESSION	BZ204853
VERSION	BZ204853.1 GI:23862905
KEYWORDS	GSS.
SOURCE	Rattus norvegicus (Norway rat)
ORGANISM	Rattus norvegicus

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 813)	Zhao, S., Shetty, J., Shatman, S., Tesgaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Riggs, F., de Jong, P. and Fraser, C.M.	Rat BAC End Sequences from Library CHORI-230 MboI segment Unpublished (1999)	Other GSSs: CH230-390D21.TJ	Contact: Shaying Zhao Department of Bukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0200

Fax: 301 838 0208
 Email: szhae@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/orering/information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_end/rat/bac_end_intro.html
 Plate: 390 row: D column: 21
 Seq primer: 77
 Class: BAC ends

FEATURES

source
 Location/Qualifiers
 1..813
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SNHsd/MCM"
 /db_xref="taxon:10116"
 /clone="CH230-390D21"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 2"
 /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
 CHORI-230 Rat (BN/SNHsd/MCM) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 7.6%; Score 34.8; DB 28; Length 813;
 Best Local Similarity 50.0%; Pred. No. 80;
 Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 178 CAAAGCGATGCGCCGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAGCGGCC 237
 548 CAAAGACCTCCACATCAACACAGACACCTCAATTAAGAAAGAAACATGAGGAAACA 607
 238 GATGAGCGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 297
 608 TCTGAGACATGAGGACCTGGAAGAAATTCCTGGAACAAACACCAATGGCTTATGCTCT 667
 298 AATGCCACCATGACCAAGTGAACGCTAAATCTCGGATATTAATGCTGCGCAA 351
 668 AAGATCAAGATCGACCAATGGATCTCAATCAATCAATCAATCAATCAATCAATCAAT 721

RESULT 7 540 bp mRNA linear EST 05-FEB-2003
 LOCUS CB210843
 DEFINITION OML01123 Oryza minuta HybridZAP-2.1 XR library Oryza minuta cDNA 5',
 mRNA sequence.
 ACCESSION CB210843
 VERSION CB210843.1 GI:28256934
 KEYWORDS EST.
 SOURCE Oryza minuta
 ORGANISM Oryza minuta
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 1 (bases 1 to 540)
 Shin, D.S.
 Oryza minuta HybridZAP-2.1 XR library
 Unpublished (2003)
 Contact: Jeong Sheop Shin
 Plant Molecular Genetics
 Graduate School of Biotechnology, University of Korea
 136-701 Anam-dong 5/1 Seoul, Korea
 Tel: 00 82 2 3290 3430
 Fax: 00 82 2 927 9028
 Email: jsshin@kucn.korea.ac.kr.
 Location/Qualifiers
 1..540
 /organism="Oryza minuta"
 /mol_type="mRNA"
 /db_xref="taxon:63629"
 /dev_stage="4-weeks after germination"

/clone_lib="Oryza minuta HybridZAP-2.1 XR library"
 /note="Organ: Immature leaf"

ORIGIN

Query Match 7.6%; Score 34.6; DB 14; Length 540;
 Best Local Similarity 47.1%; Pred. No. 77;
 Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
 225 TGTAAAGCGCGCCGATGAGGCGGATGATGATGATGATGATGATGATGATGATGATG 284
 140 TGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 199
 285 TGTGTTGAGAAATATATGACCATGACCAAGTGAACGCTAAATCTCGATATTAATGCT 344
 200 TGTGTTGAGAAATGACCAATGACCAATGACCAATGACCAATGACCAATGACCAATGAC 259
 345 CGGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 393
 260 GAGAGCTCTGCAAGACACATATCTATGATGATGATGATGATGATGATGATGATGATG 308

RESULT 8

AV976624 556 bp mRNA linear EST 14-MAR-2002
 AV976624 Nori Satoh unpublished cDNA library, egg Clona
 intestinalis cDNA clone cleg42j24 5', mRNA sequence.
 AV976624
 AV976624.1 GI:19466390
 EST.
 Clona intestinalis
 Clona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Cloniidae; Clona.

REFERENCE 1 (bases 1 to 556)
 Satoh, N., Satoh, Y., Kohara, Y. and Shin-I, T.
 Expressed genes in Clona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.

FEATURES

source
 Location/Qualifiers
 1..556
 /organism="Clona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="cleg42j24"
 /tissue_type="whole animal"
 /dev_stage="egg"
 /clone_lib="Nori Satoh unpublished cDNA library, egg"

ORIGIN

Query Match 7.6%; Score 34.6; DB 9; Length 556;
 Best Local Similarity 47.1%; Pred. No. 78;
 Matches 106; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
 222 TTATGTTAAGCGCCGATGAGGCGGATGATGATGATGATGATGATGATGATGATGATG 281
 67 TGAATCAAGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126
 282 GAATGTTTCAAGAAATATATGACCATGACCAAGTGAACGCTAAATCTCGATATTAATG 341
 127 TGAATGACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 186
 342 TGTGCGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 401
 187 TGAATGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 446
 402 CGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 446

DB	247	TGATGTCAAAGCTGATGTGTCACCTACCAATGAATGTCAACACTTA	231
RESULT 9			
LOCUS	BQ141266	539 bp	mRNA linear EST 26-APR-2002
DEFINITION	NP017G03PH1F1024 Phoma-infected Medicago truncatula cDNA clone		
ACCESSION	NP017G03PH 5', mRNA sequence.		
VERSION	BQ141266.1	GI:20277392	
KEYWORDS	EST.		
SOURCE	Medicago truncatula (barrel medic)		
ORGANISM	Medicago truncatula		
REFERENCE	1 (bases 1 to 539)		
AUTHORS	Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R., Gonzalez,R.A., Bell,C.D., Imman,D.T., Maugh,M.E., Sullivan,J.P., May,G.D. and Paiva,N.L.		
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation		
JOURNAL	Medicago truncatula Phoma-infected library		
COMMENT	Unpublished (2002)		
	Contact: Paiva NL		
	Plant Biology Division		
	The Samuel Roberts Noble Foundation		
	2510 Sam Noble Parkway, Ardmore, OK 73402, USA		
	Tel: 580 221 7317		
	Fax: 580 221 7380		
	Email: nlpaiva@noble.org		
	Insert Length: 539 Std Error: 0.00		
	Plate: 017 row: G column: 03		
	Seq primer: TCACACAGAAACAGCTATGAC.		
FEATURES	Location/Qualifiers		
source	1..539		
	/organism="Medicago truncatula"		
	/mol_type="mRNA"		
	/db_xref="taxon:3880"		
	/clone="NF017G03PH"		
	/issue_type="leaf"		
	/dev_stage="Pathogen-induced, young trifoliolate"		
	/clone_l1b="Phoma-infected"		
	/note="Vector: pBluescript SK(-); Young trifoliolate leaves of Medicago truncatula were excised and dip-inoculated in a spore suspension of Phoma medicaginis, and incubated in humid dishes. Pools of leaves were harvested at 0, 15, and 30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96 hours, and used to prepare total RNA. cDNA was prepared from polyA+ enriched, pooled samples of equivalent amounts of total RNA from each sample. The cDNA was directionally ligated into the Uni-Zap XR vector (Stratagene) and packaged using the Gigapack III Gold packaging extracts. Phagemids containing cDNA inserts were in vivo excised from the recombinant Uni-Zap XR vector using ExAssist helper phage and the E. coli strain XL1-Blue MRP' (Stratagene). Excised plasmids were plated using SOLR cells."		
ORIGIN			
Query Match	7.5%;	Score 34.4;	DB 13; Length 539;
Best Local Similarity	46.0%;	Pid. No. 88;	
Matches 116;	Conservative 0;	Matches 136;	Indels 0; Gaps 0;
129	AACGTGAGCATTTATCAGTACGGTCCGCTTAAGCGCGCTTCTCGAAAGCATGC	188	
DB	233 AATTTTAAACATCTGACTCCGGTACATTAATGATACCTTTGAATGAAGACGAGGC	292	
189	CCGTAAATCTGAAACGACATTACCCAGAGCGCTTAATGTAACGCGCCGATGTAGACCA	248	
DB	293 ACGATATCCACCTGAGTTCTCCACAAATTCGCAAAACACCTTGGTCCGCTATCTCGTAA	352	
249	GGGTGCGGATATATGTAATGTAATGTAATGTAATGTTTCAGAAATTAATGCCACCAT	308	

Db	353	TGTCGATAGCTTTTAACTTCTCTTCATTCCTCCAGAACCAATTCTAAATCAATCTTTATAT	412
Qy	309	CGACCAAGGAGACCGTAAACATCCGATATTAATCTGTCGGCCAAATATGATACAGTGGTAC	368
Db	413	TGAGAAATAGAGAGCCACCAAGAATCTGTGTGGCTTTGGGATACACTATTCGCTTTCGTAC	472
Qy	369	CCGTGTGTGTAC 380	
Db	473	CAAAATTTGTAC 484	
RESULT 10			
LOCUS	AM686303	560 bp	linear
DEFINITION	NF036D06NR1F1000 Nodulated root Medicago truncatula cDNA clone		
ACCESSION	AM686303		
VERSION	AM686303.2	GI:11931130	
KEYWORDS	EST.		
SOURCE	Medicago truncatula (barrel medic)		
ORGANISM	Medicago truncatula		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.		
AUTHORS	1 (bases 1 to 560) Watson,B.S., Shih,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores,H.R., Imman,J.T., Weller,J.W., May,G.D. and Paiva,N.L.		
TITLE	Expressed Sequence Tags from the Samuel Roberts Noble Foundation Medicago truncatula nodulated root library		
JOURNAL	Unpublished (2000)		
COMMENT	On Apr 14, 2000 this sequence version replaced gi:7561039. Contact: Paiva NL. Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73402, USA Tel: 580 221 7317 Fax: 580 221 7380 Email: nlpaiva@noble.org Insert Length: 665 Std Error: 0.00 Plate: 036 row: D column: 06 Seq primer: TCACACAGAAACGATGATAC. Location/Qualifiers 1..560 /organism="Medicago truncatula" /mol_type="mRNA" /db_xref="taxon:3880" /clone="NF036D06NR" /tissue_type="root" /dev_stage="pooled developmental" /clone_lib="Nodulated root" /note="Vector: Lambda Zap; Four-week old Rhizobium meliloti-inoculated Medicago truncatula roots, containing a mixture of young and old roots and nodules."		
FEATURES	source		
ORIGIN	Query Match 7.5%; Score 34.4; DB 10; Length 560; Best Local Similarity 46.0%; Pred. No. 89; Matches 116; Conservative 0; Mismatches 136; Indels 0; Gaps 0;		
Qy	129	AACGTAGACATTATCATGATACGGTCCGCTAACGCTGCGCTTCTCTGCAGAACGATGC	188
Db	230	AATTTTAAAAACATCTGACTCCGTTTACATATATGATACCTCTTGATATGAGACGAGC	289
Qy	189	CCGTAATCTGAAACGACATTAACCCAGAGCGTTATGTTAAGCGCGCCGATGAGCCA	248
Db	290	ACGGATTCACCTGAGATTCCTCCACAAATGGCAAAACACCTTGGTCGGATATCTCGCTAA	349
Qy	249	GGGTGCGGATATATGATATGATGAACTGATCAGAAATGTTTCAGAAATATGCAACAT	308
Db	350	TGTGATAGTATTAGTTCCTTCATATGCTCCAGAAAGCAATTCATAATCATCTTTATAT	409

QY 129 AACGTGACATTATTCAGTACGGTTCCGCTAACGCTGCGTTGCTCGAACCAGATGC 188
177 AATTTAAAAACATCTGACTCCGGTTACATTAATGCAATACCTTTGAAATGAAGACGAGGC 236
Db

[illegible]

RESULT 13					
LOCUS	BJ367372				
DEFINITION	BJ367372	524 bp	mrna	linear	EST 08-MAR-2002
ACCSSION	dictyostelium discoidium cdna library, CF				
ACCESSION	dictyostelium cdna clone ddc42p08 5', mRNA sequence.				
VERSION	BJ367372.1	GI:19276674			

FEATURES	source
KEYWORDS	EST.
SOURCE	Dictyostelium discoideum
ORGANISM	Dictyostelium discoideum
REFERENCE	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
AUTHORS	1 (bases 1 to 524)
TITLE	Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
JOURNAL	Full length cDNA of Dictyostelium discoideum at the culmination stage
COMMENT	Unpublished (2002) Contact: Tadao Shin-i Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. Location/Qualifiers
FEATURES	1..524
source	/organism="Dictyostelium discoideum" /mol_type="mRNA" /strain="AX4" /db_xref="taxon:44689" /clone="ddc42p08" /sex="mat A" /dev_stage="Culmination stage" /clone_lib="Dictyostelium discoideum cDNA library, CF"
ORIGIN	
Query Match	7.5%; Score 34.2; DB 12; Length 524;
Beat Local Similarity	55.5%; Pred. No. 99;
Matches	66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
Oy	238 GATGTAGCCAGGTCGGATATAGTACTATTGAACTGACTCAGAAATGTTTCAGAAAT 297
Db	181 GATGAAGATGATGATTAATGTAGAGATGATATATTGGAAATCAATAGAGACAAATAT 240
Oy	298 AATGCCACCATCGACCACTGAGAACCTTAAATCTCCGATATTACTGTGGCCCAATATGA 356
Db	241 AATAACACCAACACACAAATATATATATAAATAATTAATTAATTAAGATGATGATGA 299
RESULT 14	
LOCUS	BJ333958 547 bp mRNA linear EST 05-MAR-2002
DEFINITION	BJ333958 Dictyostelium discoideum cDNA library, AF Dictyostelium
ACCESSION	dictoideum cDNA clone dda4j16 5', mRNA sequence.
VERSION	BJ333958
KEYWORDS	BJ333958.1 GI:19164088
SOURCE	EST.
ORGANISM	Dictyostelium discoideum
REFERENCE	Dictyostelium discoideum
AUTHORS	Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
TITLE	1 (bases 1 to 547)
JOURNAL	Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
COMMENT	Full length cDNA of Dictyostelium discoideum at the aggregation stage Unpublished (2002) Contact: Tadao Shin-i Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. Location/Qualifiers
FEATURES	1..547
source	/organism="Dictyostelium discoideum" /mol_type="mRNA" /strain="AX4" /db_xref="taxon:44689" /clone="dda4j16" /sex="mat A" /dev_stage="Aggregation stage"

```

ORIGIN
      /clone_1lib="Dictyostelium discoideum cDNA library, Af"

Query Match      7.5%; Score 34.2; DB 12; Length 547;
Best Local Similarity .55.5%; Pred. No. 1e+02;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 238 GATGTAGGCCAGGTCGGATTAATAGTACTTGTGAACCTGACTCAGATGCTTTCAGAAAT 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 188 GATGAAGATGATGTAATTGCGAATATGATATTTTGATTCATATGAATGAGCAATATAT 247

QY 298 AATGCCACCATGACCAAGTGAACGCTTAAACTCCGATATTACTGTGGCCCAATATGA 356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 AATTAACAACAACAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 302

RESULT 15
LOCUS      BUJ336903 553 bp mRNA linear EST 05-MAR-2002
DEFINITION BUJ336903 Dictyostelium discoideum cDNA library, Af Dictyostelium
            discoideum cDNA clone dda55g13 5', mRNA sequence.
ACCESSION  BUJ336903
VERSION     BUJ336903.1 GI:19167033
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum
ORGANISM    Dictyostelium discoideum
REFERENCE    Eukaryota; Mycetozoa; Dictyostelid; Dictyostelium.
AUTHORS     1 (bases 1 to 553)
            Urushihara,H., Tanaka,Y., Kohara,Y. and Shih-I,T.
TITLE       Full length cDNA of Dictyostelium discoideum at the aggregation
            stage
JOURNAL      Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshin@gene.nig.ac.jp.
            Location/Qualifiers
                1..553
                /organism="Dictyostelium discoideum"
                /mol_type="mRNA"
                /strain="AX4"
                /db_xref="taxon:44689"
                /clone="dda55g13"
                /sex="mat A"
                /dev_stage="Aggregation stage"
                /clone_lib="Dictyostelium discoideum cDNA library, Af"

ORIGIN
      /clone_1lib="Dictyostelium discoideum cDNA library, Af"

Query Match      7.5%; Score 34.2; DB 12; Length 553;
Best Local Similarity .55.5%; Pred. No. 1e+02;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 238 GATGTAGGCCAGGTCGGATTAATAGTACTTGTGAACCTGACTCAGATGCTTTCAGAAAT 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 184 GATGAAGATGATGTAATTGCGAATATGATATTTTGATTCATATGAATGAGCAATATAT 243

QY 298 AATGCCACCATGACCAAGTGAACGCTTAAACTCCGATATTACTGTGGCCCAATATGA 356
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 AATTAACAACAACAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATGA 302

Search completed: March 16, 2004, 04:29:05
Job time : 2235.91 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgctacacagtactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenBml.*
2: gb_ba.*
3: gb_hcg.*
4: gb_in.*
5: gb_om.*
6: gb_ov.*
7: gb_pat.*
8: gb_ph.*
9: gb_pl.*
10: gb_pr.*
11: gb_ro.*
12: gb_sts.*
13: gb_sy.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vl.*
30: em_hcg_hum.*
31: em_hcg_inv.*
32: em_hcg_other.*
33: em_hcg_mus.*
34: em_hcg_pln.*
35: em_hcg_rtd.*
36: em_hcg_mam.*
37: em_hcg_vrt.*
38: em_ey.*
39: em_hlgo_hum.*
40: em_hlgo_mus.*
41: em_hlgo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	403.2	88.4	456	6 144909	144909 Sequence 58
2	403.2	88.4	2067	1 SEU43280	U43280 Salmonella
3	395.2	86.7	5103	1 STRA2301	AJ002301 Salmonella
4	395.2	86.7	22411	1 AE008749	AE008749 Salmonella
5	393.6	86.3	254050	1 AL627269	AL627269 Salmonella
6	393.6	86.3	301983	1 AE016840	AE016840 Salmonella
7	387.8	85.0	1048	1 STRA2301	AJ002301 Salmonella
8	284.2	62.3	361	6 144908	144908 Sequence 56
9	278.4	61.1	2889	1 CSP515700	AJ515700 Citrobacter
10	229.2	50.3	2920	1 CFS515701	AJ515701 Citrobacter
11	225.4	49.4	1711	1 AF275733	AF275733 Escherichia
12	225.4	49.4	10190	1 AE005315	AE005315 Escherichia
13	225.4	49.4	327773	1 AP002554	AP002554 Escherichia
14	224	49.1	4680	1 ECCSGABDG	AE002554 Escherichia
15	224	49.1	10346	1 AE000205	AE000205 Escherichia
16	224	49.1	15047	1 D90741	D90741 Escherichia
17	223.8	49.1	306358	1 AE016759	AE016759 Escherichia
18	222.4	48.8	456	6 AX814811	AX814811 Sequence
19	222.4	48.8	648	1 ECOCSSGAA	U49978 Escherichia
20	208	45.6	2883	1 ESAS15702	AJ515702 Enterobacter
21	203.2	44.6	230	1 SEU53207	U53207 Salmonella
22	186.4	40.9	10370	1 AE015131	AE015131 Shigella
23	186.4	40.9	292504	1 AE016981	AE016981 Shigella
24	163.4	35.8	437	1 AP237726	AP237726 Shigella
25	93.2	20.4	19201	1 D90742	D90742 Escherichia
26	50	11.0	1212	1 ECOL131756	AJ131756 Escherichia
27	48.2	10.6	78	6 AX814809	AX814809 Sequence
28	41.6	9.1	72	6 AX814798	AX814798 Sequence
29	36.2	7.9	1425	1 AF343445	AF343445 Lactobacillus
30	36.2	7.9	302156	1 AC116977	AC116977 Dictyostelium
31	35.6	7.8	13430	1 AE000409	AE000409 Escherichia
32	35.6	7.8	108366	2 AC020660	AC020660 Homo sapiens
33	35.6	7.8	110000	2 ECOLM67_2	Continuation (3 of AC145807 Silvana
34	35.6	7.8	110000	2 AC145807_0	AC145807 Silvana
35	35.6	7.8	196539	2 AC145912	AC145912 Gallus gallus
36	35.6	7.8	168267	2 AC146009	AC146009 Pan troglodytes
37	35.2	7.7	163619	2 AC118493	AC118493 Rattus norvegicus
38	34.6	7.6	11160	1 AE008209	AE008209 Agrobacterium
39	34.6	7.6	11476	1 AE009405	AE009405 Agrobacterium
40	34.6	7.6	125020	9 AF429315	AF429315 Homo sapiens
41	34.2	7.5	1458	1 AF173043	AF173043 Orientia
42	34.2	7.5	179906	2 AC119683	AC119683 Rattus norvegicus
43	34.2	7.5	268098	2 AC133387	AC133387 Rattus norvegicus
44	34.2	7.5	299708	2 AC110963	AC110963 Rattus norvegicus
45	34	7.5	956	8 AK058943	AK058943 Oryza sativa

ALIGNMENTS

RESULT 1
144909 LOCUS 144909 456 bp DNA
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION 144909
VERSION 144909.1 GI:2469622
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
AUTHORS Dorian,J.V., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the egfa gene for detection of Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

Pred. No. is the number of results predicted by chance to have a

FEATURES
source location/Qualifiers
1..456
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match

Best Local Similarity 88.4%; Score 403.2; DB 6; Length 456;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGAGCATTCGAGCATTCGATGTTCTGGCGTCTTGCT 60
DB 1 ATGAACCTTTTAAAGTGGAGCATTCGAGCATTCGATGTTCTGGCGTCTTGCT 60
QY 61 GCGCTTATGATCAGCTGTTACCCGCTGTTTACCATGAAATGACATGATCCGCG 60
DB 61 GCGCTTATGATCAGCTGTTACCCGCTGTTTACCATGAAATGACATGATCCGCG 60
QY 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTTCGCTAACCGCTTCTGCTGCA 120
DB 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTTCGCTAACCGCTTCTGCTGCA 120
QY 181 AGCGATGCCGTAATCTGAAACGACATTAACCGAGCGGTATGCTAAACGCGCGAT 180
DB 181 AGCGATGCCGTAATCTGAAACGACATTAACCGAGCGGTATGCTAAACGCGCGAT 180
QY 241 GTAGGCGAGGTCGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 240
DB 241 GTAGGCGAGGTCGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 240
QY 301 GGCACCATTCGACCACTGAGCACTAAACCTCCGATTAATCTGCTGGCAATACGCGGT 300
DB 301 GGCACCATTCGACCACTGAGCACTAAACCTCCGATTAATCTGCTGGCAATACGCGGT 300
QY 361 AATAACGCCGCGCTGTTAATCAAGCCGATCTGATTCAGCGTATGCTGCTAGGTT 360
DB 361 AATAACGCCGCGCTGTTAATCAAGCCGATCTGATTCAGCGTATGCTGCTAGGTT 360
QY 421 GGTTCGCAACACGCGCACTGATTAATCAAGCCGATCTGATTCAGCGTATGCTGCTAGGTT 420
DB 421 GGTTCGCAACACGCGCACTGATTAATCAAGCCGATCTGATTCAGCGTATGCTGCTAGGTT 420

```

RESULT 2
LOCUS SEU43280 2067 bp DNA linear BCT 14-FEB-1996
DEFINITION *Salmonella enteritidis* agfBAC operon: fimbria-like protein
precursor (agfB), thin aggregative fimbriae precursor (agfA), and
AgfC (agfC) genes, complete cds.
VERSION U43280
KEYWORDS U43280.1 GI:1184712
SOURCE *Salmonella enteritidis*
ORGANISM *Salmonella enteritidis*
REFERENCE 1 (bases 1 to 2067)
AUTHORS Doran, J.L., Collinson, S.K., Burian, J., Santos, G., Todd, B.C.,
Munro, C.K., Kay, C.M., Baner, P.A., Peterkin, P.I., and Kay, W.W.
TITLE DNA-based diagnostic tests for *Salmonella* species targeting agfA,
the structural gene for thin, aggregative fimbriae
JOURNAL J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
MEDLINE 94013373
PUBMED 8104955
REFERENCE 2 (bases 1 to 2067)
AUTHORS Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and
Kay, W.W.
TITLE *Salmonella enteritidis* agfBAC operon encoding thin, aggregative
fimbriae
JOURNAL J. Bacteriol. 178 (3), 662-667 (1996)
MEDLINE 96146512
PUBMED 850497
REFERENCE 3 (bases 1 to 2067)

AUTHORS

Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
Direct Submission
Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
Victoria, BC V8W 3P6, Canada

FEATURES

source

location/Qualifiers
1..2067

/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrc"
571..576
598..603
696..1151
/gene="agfB"
/gene="agfB"
696..1151
/note="fimbria-like protein precursor"
/codon_start=1
/product="agfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKKKLIFMMLTIGAPGATATNYDLASEYNFAVNELSKSPN
QALITGVGTNSARVROEGSKLSVISOEGNNRAKVDQAGNYPAYIEQGNANDA
SISQAYGNSAIIQKSGNKAVITQGTQKRAVVVQKSHAIRVQR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
/product="AgfB"
/note="putative signal sequence of 21 amino acids;
fimbria-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbria subunit precursor; major
fimbrial subunit of thin aggregative fimbriae precursor"
/codon_start=1
/product="AgfA fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLKVAFPAIVVSSALAGVVPWGCGGNGGNGSGSPDST
LSIYQGSANALALQSDARKSPRTTQSGTGGCAVGGAGDNTIELTONGFRNAT
IDQWNAKNSDITVQYGNNALVNOTASDSVVRQVFGNNATANGY"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/product="AgfA fimbria"
/note="thin aggregative fimbria subunit; major fimbrial
subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=1
/product="agfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MHTLLALASNGITPTTQGGDIYVLPQVTLNRPYQVQVLT
SYRQGVGQSTKQKQTLSPANQPIELSTVNSISSEDSVKIIVSDGSLHLSQO
WPPSAQ"

CDS

sig_peptide
mat_peptide

gene
CDS

sig_peptide
mat_peptide

stem_loop
gene

CDS

ORIGIN

Query Match
Best Local Similarity 88.4%; Score 403.2; DB 1; Length 2067;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

Qy 1 ATGAACCTTTTAAAGTGGACACATTCGACGAATCGTAGTTTCGGCAGTCTCTGGCT 60
 Db 1193 ATGAACCTTTTAAAGTGGACACATTCGACGAATCGTAGTTTCGGCAGTCTCTGGCT 1252
 Qy 61 GCGCTCTATGATCACTGCTTACCCGCTGTGTTTACCCCAATGGAATGACACATCATCCGAC 120
 Db 1253 GCGCTCTATGATCACTGCTTACCCGCTGTGTTTACCCCAATGGAATGACACATCATCCGAC 1312
 Qy 121 CCGGACTCAACGCTTACGATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 180
 Db 1313 CCGGACTCAACGCTTACGATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 1372
 Qy 181 AGCGATGCGCGGTAATCTGAAAACGACATTCACCGAGCGGTTATGATGATGATGATGATGAT 240
 Db 1373 AGCGATGCGCGGTAATCTGAAAACGACATTCACCGAGCGGTTATGATGATGATGATGATGAT 1432
 Qy 241 GTAGGCGCAGGTCGCGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 Db 1433 GTAGGCGCAGGTCGCGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1492
 Qy 301 GCCACCATCGACCGATGGAACGCTTAAACTCCGATTTACTGTCGGCCAAATACGGCGGT 360
 Db 1493 GCCACCATCGACCGATGGAACGCTTAAACTCCGATTTACTGTCGGCCAAATACGGCGGT 1552
 Qy 361 AATAACGCGCGCGCTGTTATATGACACCGCATTTGATTCAGGCTAATGTCGCTCAGGTT 420
 Db 1553 AATAACGCGCGCGCTGTTATATGACACCGCATTTGATTCAGGCTAATGTCGCTCAGGTT 1612
 Qy 421 GGTITGGCAACACGCGCAGGCTTACCAAGTATTAA 456
 Db 1613 GGTITGGCAACACGCGCAGGCTTACCAAGTATTAA 1648

/db_xref="GI:2739233"
 /db_xref="GOA:O54290"
 /db_xref="SWISS-PROT:O54290"
 /translation="CTLISGMVMAVLVIIHGSALAPHLIEIVSYLTGIVAFIMCIQAKQ
 LLISFVPGTFIGACATFAGCGDMRLVLPISLALGIIFGYAMKNSGIMLASRREHSANT
 AVTR"
 complement(361..1194)
 /gene="csgG"
 complement(361..1194)
 /gene="csgG"
 /function="lipoprotein"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="CsgG protein"
 /protein_id="CA05312.1"
 /db_xref="GI:2739234"
 /db_xref="GOA:O54291"
 /db_xref="SWISS-PROT:O54291"
 /translation="MPRLIIIVAVLLISGCTAPPKQAARPTLMPRAOSYKDLTHLPA
 PTGKIFVSVNIQDETCQFEPASNSSTAVPQSAITAVTALADSRNFTLEKQGLQ
 NLINERKIIIRAQENGTAVMNNRIPIQSLTANITMVGSIIGYSNVASGGVARYFG
 IGADTQYQLOQIAVNLRVNVSTGEIISVTSKTIISYEQAVFPFIDYQRLIEG
 IGYTSNPVMLCLMSALETGVIFLINDIGRGLMDLQNKADRDNDILVKYRELAVPPE
 S"
 complement(1221..1637)
 /gene="csgF"
 complement(1221..1637)
 /gene="csgF"
 /codon_start=1
 /transl_table=11
 /product="CsgF protein"
 /protein_id="CA05313.1"
 /db_xref="GI:2739235"
 /db_xref="SWISS-PROT:O54292"
 /translation="MRVKAHVLMLFSPPLTLAGNMTFOFRNPFGNPNNGSFLANS
 AQONSYKDPAYNDNDFGIEPSPALDNFTQAIQSOIILGGLLTNINTGKGRVNTDFII
 DIANRDQLOLANTDRKTRGRTIEVSGLOIQTSDP"
 complement(1664..2059)
 /gene="csgB"
 complement(1664..2059)
 /gene="csgB"
 /codon_start=1
 /transl_table=11
 /product="CsgB protein"
 /protein_id="CA05314.1"
 /db_xref="GI:2739236"
 /db_xref="SWISS-PROT:O54293"
 /translation="MKRYLTWIVAAEELIPATGNLANEVEYVEPGLLTDHTVSSIGHE
 FYRPFSPKSESYGNLTINERPARSGMTTIVNDVIFQTLFPMKRDPEKTIVF
 ALAQTERALNRQIDQTLTSTSLARDF"
 complement(2064..2714)
 /gene="csgD"
 complement(2064..2714)
 /gene="csgD"
 /function="transcriptional regulator"
 /codon_start=1
 /transl_table=11
 /product="CsgD protein"
 /protein_id="CA05315.1"
 /db_xref="GI:2739237"
 /db_xref="GOA:O54294"
 /db_xref="SWISS-PROT:O54294"
 /translation="MFEVHSSHGHTLLITKPSLQATVALLQHLKQSLATITKLIANIQ
 RSLSDISAGCIVLMDMMADKLIHYQDNISSKNNI KTLILNTPPDYPRLENNP
 HNGVFATDEQHEVVGLOGILGECEYFQKLSAYIITISGNTRYNSTSALATHEE
 KEILNKLRIGASNNIEARSLFISENTVKTHLYNLFKIAVGNRTQAVSWANDLNR"
 complement(2869..2896)
 /gene="csgDBFG"
 complement(2869..2896)
 /gene="csgDBFG"
 complement(2891..2896)
 /gene="csgDBFG"

RESULT 3
 STAJ2301
 LOCUS
 DEFINITION Salmomella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and k
 csgC genes.
 ACCESSION AJ002301
 VERSION AJ002301.1
 KEYWORDS
 SOURCE
 ORGANISM
 Salmomella typhimurium
 Salmomella typhimurium
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmomella.
 REFERENCE
 AUTHORS Romling U., Bian Z., Hammar M., Sierralta M.D. and Normark S.
 TITLE Cui1 fibers are highly conserved between Salmomella typhimurium
 and Escherichia coli with respect to operon structure and
 regulation
 JOURNAL J. Bacteriol. 180 (3), 722-731 (1998)
 MEDLINE 98117058
 PUBMED 9457880
 REFERENCE
 AUTHORS Romling U.
 TITLE Direct Submmission
 JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
 Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
 FEATURES
 source
 1..5103
 /organism="Salmomella typhimurium"
 /mol_type="genomic DNA"
 /strain="SR-11"
 /db_xref="taxon:602"
 <1..332
 /note="ORF179a"
 /codon_start=2
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="CA05311.1"

gene
 CDS
 gene
 CDS
 gene
 -10_signal
 -35_signal

sig_peptide	mat_peptide	gene	CDS
3344. .3376	/gene="csgba"	-35_signal	3344. .3349
	/gene="csgba"	-10_signal	3371. .3376
	/gene="csgba"	gene	3469. .3924
	/gene="csbg"	CDS	3469. .3924
	/function="nucleator"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="Csgb protein"		
	/protein_id="CAO5316.1"		
	/db_xref="GI:2739238"		
	/db_xref="GOA:P55226"		
	/translation="MKKLLFMWLTTLTLLAIPATATNYDLARSEYFAVNEISKSRRPDA QALISGAVGQDGNAPRQSGSLVISEGNNRAKVDQAGNPAVTEGQGNANDA SISQAVGSAALITQGSNNKANITQYGTOKTAVVQKSHAIRTOR"		
3469. .3531	/gene="csbg"		
	/gene="csbg"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="Csga protein"		
	/function="nucleator"		
	/protein_id="CAO5317.1"		
	/db_xref="GI:2739239"		
	/db_xref="GOA:P55225"		
	/translation="MKLLKVAFAIVVSGSALAGVVPQWGGGNGNGSGSDST LSIYQGSANALALQSDARKSETTIVSGYGNAGADVQAGADSTTELTONGPNNAT IDQNNANNSITVQYGGNNALVNOTASDSVWVQYGFNNATANY"		
3966. .4025	/gene="csga"		
	/gene="csga"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="Csga protein"		
	/protein_id="CAO5317.1"		
	/db_xref="GI:2739239"		
	/db_xref="GOA:P55225"		
	/translation="MKLLKVAFAIVVSGSALAGVVPQWGGGNGNGSGSDST LSIYQGSANALALQSDARKSETTIVSGYGNAGADVQAGADSTTELTONGPNNAT IDQNNANNSITVQYGGNNALVNOTASDSVWVQYGFNNATANY"		
4026. .4418	/gene="csga"		
	/product="Csga protein"		
	/function="fiber subunit"		
	/protein_id="CAO5317.1"		
	/db_xref="GI:2739239"		
	/db_xref="GOA:P55225"		
	/translation="MKLLKVAFAIVVSGSALAGVVPQWGGGNGNGSGSDST LSIYQGSANALALQSDARKSETTIVSGYGNAGADVQAGADSTTELTONGPNNAT IDQNNANNSITVQYGGNNALVNOTASDSVWVQYGFNNATANY"		
4483. .4809	/gene="csbg"		
	/gene="csbg"		
	/codon_start=1		
	/transl_table=11		
	/product="Csgc protein"		
	/protein_id="CAO5318.1"		
	/db_xref="GI:2739240"		
	/translation="MHTLLILALASNOITFTTQOQDITVIVIVQTLNFCVCQVQIL SYRDSVGQSHSTQOKTSLIPANQPIELSRSLVSISSDVKITVVSQSLHSQO WPSAQ"		
4483. .4533	/gene="csbg"		
	/gene="csbg"		
	/codon_start=1		
	/transl_table=11		
	/product="Csgc protein"		
	/protein_id="CAO5318.1"		
	/db_xref="GI:2739240"		
	/translation="MHTLLILALASNOITFTTQOQDITVIVIVQTLNFCVCQVQIL SYRDSVGQSHSTQOKTSLIPANQPIELSRSLVSISSDVKITVVSQSLHSQO WPSAQ"		
4534. .4806	/gene="csbg"		
	/gene="csbg"		
	/codon_start=1		
	/transl_table=11		
	/product="Csgc protein"		
	/protein_id="CAO5318.1"		
	/db_xref="GI:2739240"		
	/translation="MHTLLILALASNOITFTTQOQDITVIVIVQTLNFCVCQVQIL SYRDSVGQSHSTQOKTSLIPANQPIELSRSLVSISSDVKITVVSQSLHSQO WPSAQ"		
4940. .5103	/gene="ORF103"		
	/gene="ORF103"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="Csgc protein"		
	/protein_id="CAO5318.1"		
	/db_xref="GI:2739240"		
	/translation="MHTLLILALASNOITFTTQOQDITVIVIVQTLNFCVCQVQIL SYRDSVGQSHSTQOKTSLIPANQPIELSRSLVSISSDVKITVVSQSLHSQO WPSAQ"		
4940. .5103	/gene="ORF103"		
	/gene="ORF103"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="Csgc protein"		
	/protein_id="CAO5318.1"		
	/db_xref="GI:2739240"		
	/translation="MHTLLILALASNOITFTTQOQDITVIVIVQTLNFCVCQVQIL SYRDSVGQSHSTQOKTSLIPANQPIELSRSLVSISSDVKITVVSQSLHSQO WPSAQ"		
4940. .5103	/gene="ORF103"		
	/gene="ORF103"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="Csgc protein"		
	/protein_id="CAO5318.1"		
	/db_xref="GI:2739240"		
	/translation="MHTLLILALASNOITFTTQOQDITVIVIVQTLNFCVCQVQIL SYRDSVGQSHSTQOKTSLIPANQPIELSRSLVSISSDVKITVVSQSLHSQO WPS		

```

/crensl_table=1
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:054295"
/translation="MSVKKKNTPAIGLCIAFFHSAVGQQTGCVHFRGALVEPL
CDISTHAENT"

```

Query Match	86.7%;	Score 395.2;	DB 1;	length 5103
Best Local Similarity	91.7%;	Pred. No. 9.1e-109;		
Matches 418; Conservative	0;	Mismatches 35		

[illegible]

Db 3966 ATGAACCTTTTAAAGTGGCAGCATTTCGAGCAATCGTAGTTCTGGCAGTGCCTCGCT 60
 0y 61 GCGGCTATGATCAGCTCCGTTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4025

D_B
OY

4026 GGGCTGTTCCACATGGGGCGCGGTATTCAATAAGCGGGGCCAATGTTCCGGC 120
121 CCGACTCAA GTTGAACA TTTTA TTAA TA AAAAA AAAAAA AAAGA AGAGT GGAGT GAAGA 4085

Ddb
4086 CCGGATTCCAGCTTGAGCAATTATCAGTAGCGTTCCGCTAACGCACGGTTCGCAA 180
2y 181 AGCGATGCCCGTAATCTGAACCGACCGAATGCGGTGCGCTGCTGCAA 4145

[illegible]

4206 GTAGGCCAGGCTGCCGATAACACTACTATTGAACTGACTCAGATCGTTTCAGAAACAAT 300

4266 GCCACCATGCACACAGTGAACGCTAAAACTCGATATTACTGTGGTCAATACGGCGGT 4325

4326 AATTAAGCGCGCGTGGTTATCAGACCGCATCTGATTCAGGCGTAAATGGTGGCTCAGGTT 4385

4386 GGTITGGCAACAAGCGACGGCTACCAAGTATTAA 4421

CUS					
008749	AE008749	22411 bp	DNA	linear	BCT 23-APR-2003
FINITION	<i>Salmonella typhimurium</i>	LT2, section 53 of 220 of 41			
	Genome				

AE008749 AE006468
AE008749.1 GI:16419641

1 (b) *Salmonella typhimurium* LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; *Salmonella*.

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latreille, P., Courtney, L., Portwoll, S., All, J., Dante, M.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K.,
Grewal, N., ...

Stoneking, T., Nhan, M., Waterston, R. and Wilson, R.K.
Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2
Nature 413 (2001) 516-522

2 (bases 1 to 22411)

Submitted (29-MAR-2001) Genome Sequencing Center, Department of

CDS

6805..7497
 /gene="STM1131"
 /note="similar to E. coli orf, hypothetical protein
 (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%

identity in aa 4 - 239"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AA120062.1"
 /db_xref="GI:16419647"
 /translation="WKINKYLLQVSEFIASSYLQATLDYRHEADTRINKRIAI
 IEKIPNGIIVDASVSGVDEODKHSIDYANAIEIGVSYINKVDYLDPGFI
 FESGPDISIKPYLRQYNDSEVYAGRYDARKNTASDSEKINPFTYGYF
 DELKLEYNFTWMSDOIKEFNKKTINEHVALLAKNLSFTPYVEGVAVRNNTDER
 QTRYRGLQVHF"
 /7668..9060
 /gene="STM1132"
 /7668..7773
 /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
 /7780..9060
 /gene="STM1132"
 /note="similar to E. coli putative transport protein
 (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%

identity in aa 25 - 418"
 /codon_start=1
 /transl_table=11
 /product="putative sugar transport protein"

Query Match 86.7%; Score 395.2; DB 1; Length 22411;
 Best Local Similarity 91.7%; Pred. No. 1e-108;
 Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATGTTGTTCTGGCAGTCTGCTGCT 60
 DB 17769 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATGTTGTTCTGGCAGTCTGCT 60
 QY 61 GGCCTATATGATCAGCTGGTTACCCGTGTTGTTACCCAGAAATGCGACATGCTCCGCC 120
 DB 17829 GGCCTATATGATCAGCTGGTTACCCGTGTTGTTACCCAGAAATGCGACATGCTCCGCC 120
 QY 121 CCGGACTCAAGTGTGAGCATTTATCAGTACGTTCCGTTACCGTGGCTTCTGTGCA 180
 DB 17889 CCGGACTCAAGTGTGAGCATTTATCAGTACGTTCCGTTACCGTGGCTTCTGTGCA 180
 QY 181 AGCGATGCCCGTAAATCTGAAACGACATTACCCAGACGGTTATGTTACGCGCGCAT 17948
 DB 17949 AGCGATGCCCGTAAATCTGAAACGACATTACCCAGACGGTTATGTTACGCGCGCAT 240
 QY 241 GTAGCGCAGGGTGGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 18008
 DB 18009 GTAGCGCAGGGTGGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 QY 301 GCCACCATCGACAGTGTGAAAGCTTAAATCCGATATTACTGTCGGCAATACGGCGGT 18068
 DB 18069 GCCACCATCGACAGTGTGAAAGCTTAAATCCGATATTACTGTCGGCAATACGGCGGT 360
 QY 361 AATAACGGCGGCTGTTAATCAGACCGCATCTGATTCAGAGTATGTTGCTCAGGTT 18128
 DB 18129 AATAACGGCGGCTGTTAATCAGACCGCATCTGATTCAGAGTATGTTGCTCAGGTT 420
 QY 421 GGTTCGCAACACGCGCAGCGCTTACCGATTTAA 456
 DB 18189 GGTTCGCAACACGCGCAGCGCTTACCGATTTAA 18224

RESULT 5
 AL627269 254050 bp DNA linear BCT 04-JUL-2003
 LOCUS Complete chromosome, serovar Typhi (Salmonella typhi) strain CT18,
 ACCESSION AL627269 AL513382
 VERSION AL627269.1 GI:16502231

KEYWORDS

Salmonella enterica subsp. enterica serovar Typhi

SOURCE

Salmonella enterica subsp. enterica serovar Typhi

ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

REFERENCE

1 (bases 1 to 254050)
 Parikh, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,

TITLE

Complete genome sequence of a multiple drug resistant Salmonella

JOURNAL

Nature 413 (6858), 848-852 (2001)

MEDLINE

21534947

PUBMED

11677608

AUTHORS

2 (bases 1 to 254050)
 Parikh, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,

COMMENT

Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
 sequencing team Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, UK
 E-mail: parikhj@sanger.ac.uk

FEATURES

Details of S. Typhi sequencing at the Sanger Centre are available
 on the World Wide Web.
 (URL, http://www.sanger.ac.uk/Projects/S_Typhi/).

source

1..254050
 location/Qualifiers

gene

/organism="Salmonella enterica subsp. enterica serovar
 Typhi"
 /mol_type="genomic DNA"
 /strain="CT18"
 /db_xref="taxon:90370"

CDS

181..2793
 /gene="STY1078"
 /note="synonym: pepN"
 181..2793
 /gene="STY1078"
 /note="Orthologue of E. coli pepN (AMPN ECOLI); Fasta hit
 to AMPN ECOLI (869 aa), 94% identity in 869 aa overlap"
 /codon_start=1
 /transl_table=11
 /product="aminopeptidase N"
 /protein_id="CAD08183.1"
 /db_xref="GI:16502232"
 /db_xref="GOA:O82770"
 /db_xref="SPRMBL:O82770"

misc_feature

532..1332
 /gene="STY1078"
 /note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
 family M1, score 245.10, E-value 1.9e-84"

misc_feature

1060..1089
 /gene="STY1078"

/note="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit
to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q82759"
/db_xref="SWISS-PROT:Q82759"
/translation="MYPRVKALFQIDPRAHEFTFOQLRITGTPTLEALVROKVP
KPTCMGLTFNPLGLAAGLDKDEICDALAMGFGSLGIVTPRPQPGNDKPLFR
LVDAEGLINRGNFNNLVNENKKAHFGIIGINRKNKOTPEVGRKDYLCME
KVYAVAGIYAINISSPNTPGRLTOYGDALDLDLTAIKNKNDLQVTHKRVPAVKI
APDLCEELIOVADSLRHNIIDVIAITTTDRSLVQCMKRCOOTGSLGRPLDKST
EIRRLSELKQPLIIGVGIDSVIAREKIAAGATLVQISGPIFGPLKEIYV
HI"
3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODHase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTRMBL:Q82758"
/translation="IMRIKIPDDNRWYDEEHDMMLDLANGMLFRSFRSKMLTPDA
FCPTGCVDDALYFSEKCRDELTGEORALVNLNLAIVRIYKLPOMPSMFFVAH
GEMWTPGMDASVWLSDTAEOVNLVVEPENNALCLLAPGVVINGRTMQLDALIK
IMNDRLKPOVHCHSFSLQAV"
complement(4716..5825)
/gene="STY1081"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82757"
/db_xref="SPTRMBL:Q82757"
/translation="MATLSRLFIHPVSMGIGLTHALADISGLAPDRIFMWESDGT
FTRAPQPMWRFTPSPLHDGLHLAPDGSALVRFDPDAPTVMGHNFARVA
PTAINQWLSGFSFSDVOLCWGPQLTARRVKHNAVPLGADGYLLTNESASLDGO
RCPAGVMEORPMLVYSGVAAMEDSMKVRLIGDVI EDVVKPSCSLFTTVSEKCO
KHSGEPLATQAFRTNODGNDVFCGNLIARNSGVIRVGDVEIILATAKAYGATT
LDBSVTPKRPDSTVITDQSGITCGNNQVLLSENOGIRIPIYSCRAGICGCCRIR
LLEGEVPLKKSAGMDGTILSCSVKTAIRLEN"
complement(4737..4955)
/gene="STY1081"
misc_feature

/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical
protein H10116 H10116 SW:YCBY_HAEIN (P44524; P43945) Faeta
scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82756"
/db_xref="SPTRMBL:Q82756"
/translation="MNSLPASTAGKEILLKIELEKAVGCVVGVHROGTRLI
YQSLMRSRLSRILPMGECKVSDLDLYEVQVINTETIFNPATRAVHFSGLNDTI
RNSQYGMKVKDAIVDAFTRKNLPRPVNDRESPLRLVNLNKTASIALDLSGDLH
LRGYRDRITGLAPIKETLAAAIYMSGQPGTPLDPMCGSGTLLIEMAWATDRAPIGL
RHSWMGFSGWAQDEAIWQEKVAKQRAKGLAEYSSHFYGSDDARVIRASNNAR
RAGIGELITFEVVDVAQNSPLPKGPVTSNPYGBRLSDPALIAHSLGRTYK
NORGKMLSLFSAASPDLSGLADQFPAKNGPLDVCQKNTIATITADSKRATVA
EDVANRIRKULKLEKVARQGIETVLDLDEIVNADRYDMVATQIAPKTV
DAQVAKRQLPDIIATLSVGIIPNNKILKTRERQKNDYQKSEGEFLEVEYNA
RLWNLITDYDITGLFLDHRILARMLGMSKQPLANFSYGSASVHAGLGAGANTTT
VDMSTRYLEWAERMLRLNGLSGRAHRLIQDQCLMLEANEQPLIFIDPPTFNSKR
MESEFDVORHVALMKDLKRLKRLKRGITIMSNNGRPRMDLEGLAEIGLTAEITQKT
LSPDFARNROIHNKMLIRAA"
5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJUK_ECOLI (554 aa), 34% identity in 524 aa
overlap

Query Match 86.3%; Score 393.6; DB 1; Length 254050;
Best Local Similarity 91.4%; Pred. No. 3.9e-108;
Matches 417; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy	1	ATGAACTTTTAAAGTGGAGCATTCGACGATCGATGTTTTCGACGCTCGGCT	60
Db	88306	ATGAACTTTTAAAGTGGAGCATTCGACGATCGATGTTTTCGACGCTCGGCT	88365
Qy	61	GGCGCTATGATCAGCTGTACCCGCTGTTTACCCATGAATGACATGATCCGGC	120
Db	88966	GGCGCTGTTTCAACATGAGGGCGGGCGGTAATCATAGAGCGGCAATAGTTCCGGG	89025
Qy	121	CCGACATCAACGTTGAGCATTTTATCATGATGAGTTCCGCTAACGTCGGCTTCTGCA	180
Db	89026	CCGATTCACACGTTGACATTTTATCATGATGAGTTCCGCTAACGTCGGCTTCTGCA	89085
Qy	181	AGGATCCCGGTAATGATGAAGACATTAACGAGCGGTTATGTTAAGCGGCGCAT	240
Db	89086	AGGATCCCGGTAATGATGAAGACATTAACGAGCGGTTATGTTAAGCGGCGCAT	89145
Qy	241	GTAAGCCAGGCGGATTAAGTACTATTGAATGAACTGACTGAGATGTTTCAGAAATAT	300
Db	89146	GTAAGCCAGGCGGATTAAGTACTATTGAATGAACTGACTGAGATGTTTCAGAAATAT	89205

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

1. *Enterobacteriaceae*; *Salmonella*.
(bases 1 to 301983)
Deng, W., Liou, S. R., Plunkett III, G., Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
Comparative Genomics of *Salmonella enterica* Serovar Typhimurium
T2 and CT18
J. Bacteriol. 185 (7), 2350-2337 (2003)
1264504
2 (bases 1 to 301983)
Deng, W., Liou, S. R., Plunkett, G., III, Mayhew, G. F., Rose, D. J.,
Burland, V., Kodoyianni, V., Schwartz, D. C. and Blattner, F. R.
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1. .301983

```

Typhi Ty2" Salmonella enterica subsp. enterica serovar
/ mol_type="genomic DNA"
/ strain="Ty2"
/ serovar="Typhi"
/ sub_species="enterica"
/ db_xref="taxon:209261"
complement(844..622)
/gene="trnpA"
/ locus_tag="t1735"
complement(164..622)
/gene="trnpA"
/ locus_tag="t1735"
/ note="corresponds to STY1225 from Accession AL513382"
/codon_start=1
/transl_table=11
/product="transposase for insertion sequence element
IS200"
/protein_id="AA069359.1"
/ db_xref="ngi:29137798"
/ translation="MGDSKLAHTRNCKYHIVAPKRYROAFGKRRVAGSI
CEMKVRLLEACCAHHMLLEIPKMSVSFMYLKGKSLALGYCFDLNF
REFKRGVYVDYGYKATKIDYIKHQLSEDMGSLPIFGSPFTGRX"
complement(844..1797)
/gene="f1gJ"
/ locus_tag="t1736"
complement(844..1797)
/gene="f1gJ"
/ locus_tag="t1736"
/ note="corresponds to STY1223 from Accession AL513382:

```

Salmonella typhi CT18"
/transf_table=11
/codon_start=1
/product="flagellar hook-associated protein 3"
/protein_id="AA069360.1"
/db_xref="GI:29137799"
/translation="MRISTOMVEEONMSGITNSQAMKLGESMGSTGRVTNPSPDDEI
AASDAVAVLPOAOGISOYALATRTLRKVSLEESVLSQTTAIQTQAKRIYVANGSTLI
SDDPRAVLTDLIDLOAGIRDLQMLNLANSIDNGSRVYFAQYKTLFAPDQAGYVAGKSEV
TVGGVDAVWVGHGTGAQTPNSTISGNVSPDSDSEKFLFMDLTALAAIKPVEEN
DYEKERAAAIIDKTRNGKSLNNVLIVRAELTQISELTSDLSGSRALGQRLQMS
NLVDVNMVSIVISYMOOALQASVYKTFIDMQMSLFGQNR"
/complement(1812..3473)
/gene="fliGK"
CDS
/locus_tag="ct1737"
complement(1812..3473)

gene
Salmonella typhi CT18
corresponds to STY1222 from Accession AL513382:
/locus_tag="ct1738"
/gene="fliG"
/locus_tag="ct1738"
/note="corresponds to STY1221 from Accession AL513382:
/codon_start=1
/transl_table=1
/product="flagellar protein FlgJ"
/protein_id="AA069362.1"
/db_xref="GI:29137801"
/translation="MIGDGLKLSAAMPDASINELTKKAGQDPANIRFVAOYEGMGM
TMPDDPDPVPLKFLSEFNYSQNALTLQLRRAIRPTPSDAPLGSQSDFLARLS
LPALASBOGQHHILALQALBSGMRGQRIRENGEFSYVFGKYLAKMKGPIRE
ITTEVNGEAKKVKAKRRVTSYLEALSYVALILRNRPYLAVTTAATAGCAVALQ
NAGAAIDPYNARKLAKSMTQDLKAMSEKSVTSYANLNLNF"
/complement(4488..5591)
/gene="fliG"
/locus_tag="ct1739"
complement(4488..5591)
/gene="fliG"
/locus_tag="ct1739"
/note="corresponds to STY1220 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="flagellar P-ring protein precursor"
/protein_id="AA069363.1"
/db_xref="GI:29137802"
/translation="MVFYKLAGIYVALYATLAAERIRDLTSVOGVRENSLIGLY
VLDLGGQITQTPPTQITLNNLSQGLIYPTGTNNQINVAWVATSGPPRQG
QITDIVVSNMGNKSLRGGTLMLPLKAVDSQVALAQNGNIVLGGAGASGSSPARG
OLNGRITVNSAKITREILPFGAGNTINILQNLNEDPMAQIIDAIINRARGGSAITL
DARTVVRVNSGNSQFVLADIQNMENVNTPQDAKVINSRGSGYNNRKTLLSDCA
VAQNLISVTNRLQNLNQNPTFFGGQIVTVPTQIDLRSGSGSLQSVASSNSLNISV

gene
Salmonella typhi CT18
corresponds to STY1222 from Accession AL513382:
/locus_tag="ct1738"
/gene="fliG"
/locus_tag="ct1738"
/note="corresponds to STY1221 from Accession AL513382:
/codon_start=1
/transl_table=1
/product="flagellar hook-associated protein 1"
/protein_id="AA069361.1"
/db_xref="GI:29137800"
/translation="MSGLINHAMSGLNAAQALNTVSNINNNVAGYTRQTTILIA
NSTLGAQMGNGYVSVYORVDAFTINOLRGKQNSGLFTREYQSGKIDNLLA
SSLSGSGSFTSLQTLVLSNAEDPAARQALIRKAGLNDQKRTDQVLRPODQV
ALGSSVAQIINNAKQIINLNDQIRMTGVAGASPNLIDQPDQIVSLNKIVAEV
VQDGTGTLNMANGYTLVQGSTARQALAAVPSADPIRTTVAVDEAGNINIEPKL
TSGISGLILFRSQDDQDRTNLTGQALAPASADPHRTGVDYGTQDGNKGFESGL
LEIDGKIVTGTGAQKNDSEFLIKRVSNALIDRNKVTNAELTAAESKLDPDVDTG
SDNRNQALDILQNSNVYAGMKPTNDAITLVSDGNKTSITLKTSITQANVAVKQL
QQQSVGVALDEEYGNLQRYQVYTLAAVQVLTANALPDALINIR"
/complement(3538..4488)
/gene="fliG"
/locus_tag="ct1738"
/gene="fliG"
/locus_tag="ct1738"
/note="corresponds to STY1221 from Accession AL513382:
/codon_start=1
/transl_table=1
/product="flagellar protein FlgJ"
/protein_id="AA069362.1"
/db_xref="GI:29137801"
/translation="MIGDGLKLSAAMPDASINELTKKAGQDPANIRFVAOYEGMGM
TMPDDPDPVPLKFLSEFNYSQNALTLQLRRAIRPTPSDAPLGSQSDFLARLS
LPALASBOGQHHILALQALBSGMRGQRIRENGEFSYVFGKYLAKMKGPIRE
ITTEVNGEAKKVKAKRRVTSYLEALSYVALILRNRPYLAVTTAATAGCAVALQ
NAGAAIDPYNARKLAKSMTQDLKAMSEKSVTSYANLNLNF"
/complement(4488..5591)
/gene="fliG"
/locus_tag="ct1739"
complement(4488..5591)
/gene="fliG"
/locus_tag="ct1739"
/note="corresponds to STY1220 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="flagellar P-ring protein precursor"
/protein_id="AA069363.1"
/db_xref="GI:29137802"
/translation="MVFYKLAGIYVALYATLAAERIRDLTSVOGVRENSLIGLY
VLDLGGQITQTPPTQITLNNLSQGLIYPTGTNNQINVAWVATSGPPRQG
QITDIVVSNMGNKSLRGGTLMLPLKAVDSQVALAQNGNIVLGGAGASGSSPARG
OLNGRITVNSAKITREILPFGAGNTINILQNLNEDPMAQIIDAIINRARGGSAITL
DARTVVRVNSGNSQFVLADIQNMENVNTPQDAKVINSRGSGYNNRKTLLSDCA
VAQNLISVTNRLQNLNQNPTFFGGQIVTVPTQIDLRSGSGSLQSVASSNSLNISV

```
gene
complement(5597..6295)
/gene="flgH"
/locus_tag="t1740"
complement(5597..6295)
/gene="flgH"
/locus_tag="t1740"
/notes="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MQKALAHYPVMAALMTLTCGAWIPAKPIVQAGTTAQPIEGY
PVANGSIFOSQPINYGTOPLEDRRPNITDITLTVLQENVASKSSAASRDGKT
SFGDTPRYLQGLFGNSRADMEASGGSFNGKAGNANFTSGTLTVTVQVANGN
LHVVEKQIAINQCTEFIRFSGVNPRTISGNSVPSQVDARIIEVYNGYINDAQN
MGMLRFFPLNSPM"
complement(6350..7132)
/gene="flgG"
/locus_tag="t1741"
complement(6350..7132)
/gene="flgG"
/locus_tag="t1741"
/notes="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flgG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISSLRIRAKTGGLDAQCTNMVIANNLAVSTNGEKORAVFEDL
LYQTRKPGASSTQTLPSGLQIGTVRPATERLSQGLSQTNNSKDAIKGQGR
FOVWLPGSTAYSTEDGSGFOVDNGQLYAGAFQVQPAITIPANALSTIGRDVSVT
QOQCAAVQVQQLNLTTFMNDTGLSIGENLYETSSGAPNESTPGINAGLHYQY
VETSNVAEELVNMIOVRAYEINSAVSTTDQMLQKLTQL"
complement(7146..7901)
/gene="flgF"
/locus_tag="t1742"
complement(7146..7901)
/gene="flgF"
/locus_tag="t1742"
/notes="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHAIITAMGASQTLNQAVTNSNLASTPGFRQALMLRAV
PVDLSLATRTLVASTPGADMTQQLDYTSRPDLVALQODGMVLVQAQAEQYTN
GNIVQPTQGLTIQGHPIGEGEITVPEGEIRIADGITSALNPGSPPTVAVER
LKVYKAGNEVQRSDGGLFRLTAQAQERLADLSIRMSGVLSGNVKEVEMAT
DMINARFEMQKRVITSVDENGRANQLMSMS"
complement(7922..9133)
/gene="flgE"
/locus_tag="t1743"
complement(7922..9133)
/gene="flgE"
/locus_tag="t1743"
/notes="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFSQAVSGMAATINIDVIGNIANSATYGFKSGTASPADMA
GSKVGLGVKAGITQDFTDGTITNGRLDVALISONGFRLVDSNGSVFYSRNQFTL

gene
complement(5597..6295)
/gene="flgH"
/locus_tag="t1740"
complement(5597..6295)
/gene="flgH"
/locus_tag="t1740"
/notes="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MQKALAHYPVMAALMTLTCGAWIPAKPIVQAGTTAQPIEGY
PVANGSIFOSQPINYGTOPLEDRRPNITDITLTVLQENVASKSSAASRDGKT
SFGDTPRYLQGLFGNSRADMEASGGSFNGKAGNANFTSGTLTVTVQVANGN
LHVVEKQIAINQCTEFIRFSGVNPRTISGNSVPSQVDARIIEVYNGYINDAQN
MGMLRFFPLNSPM"
complement(6350..7132)
/gene="flgG"
/locus_tag="t1741"
complement(6350..7132)
/gene="flgG"
/locus_tag="t1741"
/notes="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flgG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISSLRIRAKTGGLDAQCTNMVIANNLAVSTNGEKORAVFEDL
LYQTRKPGASSTQTLPSGLQIGTVRPATERLSQGLSQTNNSKDAIKGQGR
FOVWLPGSTAYSTEDGSGFOVDNGQLYAGAFQVQPAITIPANALSTIGRDVSVT
QOQCAAVQVQQLNLTTFMNDTGLSIGENLYETSSGAPNESTPGINAGLHYQY
VETSNVAEELVNMIOVRAYEINSAVSTTDQMLQKLTQL"
complement(7146..7901)
/gene="flgF"
/locus_tag="t1742"
complement(7146..7901)
/gene="flgF"
/locus_tag="t1742"
/notes="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHAIITAMGASQTLNQAVTNSNLASTPGFRQALMLRAV
PVDLSLATRTLVASTPGADMTQQLDYTSRPDLVALQODGMVLVQAQAEQYTN
GNIVQPTQGLTIQGHPIGEGEITVPEGEIRIADGITSALNPGSPPTVAVER
LKVYKAGNEVQRSDGGLFRLTAQAQERLADLSIRMSGVLSGNVKEVEMAT
DMINARFEMQKRVITSVDENGRANQLMSMS"
complement(7922..9133)
/gene="flgE"
/locus_tag="t1743"
complement(7922..9133)
/gene="flgE"
/locus_tag="t1743"
/notes="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFSQAVSGMAATINIDVIGNIANSATYGFKSGTASPADMA
GSKVGLGVKAGITQDFTDGTITNGRLDVALISONGFRLVDSNGSVFYSRNQFTL

Query Match 86.3%; Score 393.6; DB 1; Length 301983;
Best Local Similarity 91.4%; Pred. No. 4e-108;
```

```
Matches 417; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 ATGAACTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTTCGCGAGTCTGCGT 60
Db 37310 ATGAACTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTTCGCGAGTCTGCGT 37351
QY 61 GGGCTTATGATCGAGCTGTACCCGCTGTTTACCAATGCAATGCGCATTCGCGC 120
Db 37350 GGGCTTATGATCGAGCTGTACCCGCTGTTTACCAATGCAATGCGCATTCGCGC 37391
QY 121 CCGACTCAACGTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCGCTTCTGCAA 180
Db 37390 CCGACTCAACGTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCGCTTCTGCAA 37431
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTTACGCGCGCAT 240
Db 37430 AGCGATGCGCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTTACGCGCGCAT 37071
QY 241 GTAGCGCAGGGTGGGATTAATAGTACTTATGAACTGACCTGAATGGTTTCAGAAATAT 300
Db 37070 GTAGCGCAGGGTGGGATTAATAGTACTTATGAACTGACCTGAATGGTTTCAGAAATAT 37011
QY 301 GCCACATCGACGACGAGTGAACGCTAAAACTCCGATATTACTGCGCCATACGCGCGT 360
Db 37010 GCCACATCGACGACGAGTGAACGCTAAAACTCCGATATTACTGCGCCATACGCGCGT 36951
QY 361 AATAACGCGCGCTGTTTAAATCAGACCGCATCTGATTCAGCGTAAATGTCGTCAGTT 420
Db 36950 AATAACGCGCGCTGTTTAAATCAGACCGCATCTGATTCAGCGTAAATGTCGTCAGTT 36891
QY 421 GGTTCGCAACAGCGCAGCGCTAACCGATTA 456
Db 36890 GGTTCGCAACAGCGCAGCGCTAACCGATTA 36855

RESULT 7
STAGEBA
LOCUS STRAGEBA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AU000514
VERSION AU000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1
REFERENCE
AUTHORS Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
TITLE Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE
AUTHORS Sukupolvi,S.S.
TITLE Direct Submersion
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kilnamy1lynkatu, 20520, FINLAND
FEATURES
source
1..1048
1./organism="Salmonella typhimurium"
2./mol_type="genomic DNA"
3./db_xref="taxon:602"
4..553
5./gene="agfB"
6./gene="agfB"
7./gene="agfB"
8./codon_start=1
9./transl_table=11
10./product="agfB protein"
11./protein_id="CA04150.1"
```

```

/db_xref="GI:2275120"
/translation="MKKLLFMMLTLLLEPIATNYDLARSEYNFAVNLSSKSPNO
AATIGVGTDSNARVHEGSKLSVLSQERGNRRARVDAGNRPAYITQCTGNADA
SISQAYGNSAISAIIOKSGSNKANTYGTCTKAVVVOKSHMAIQANTIOYGTG
KTAIVVVOKSHMAIRVTR"
593..1048
/gene="agfa"
593..1048
/cds
593..1048
/gene="agfa"
/codon_start=1
/translation="agfa protein"
/protein_id="CA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:033802"
/translation="MKLLKVAAPALIVSASAVGVVPOMGGGNNHNGSSGPDST
LSIYQGSANALALOSDARKSETTITGSGVGNADVGAGDINSITELTONGFRNNAT
IDONNAKNSDITVGOYGNNAALVNOIASDSVVRQVGFNNAPANQYN"

```

Query Match 85.0%; Score 387.8; DB 1; Length 1048;
 Best Local Similarity 90.8%; Pred. No. 1.4e-106;
 Matches 413; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACATCGTAGTTCTGCGAGTCTGAGCT 60
DB 593 ATGAACTTTTAAAGTGGCAGCATTCGACATCGTAGTTCTGCGAGTCTGAGCT 60
QY 61 GGGCTATGATGAGTGTGTTACCGGTGTTTCCCAATGCAATGCAATGCAATGCGGC 120
DB 653 GGGCTGTTCCCAATGAGGGCGCGGCTTAATCAATACGGGCGGCAATGTTGCGGC 712
QY 121 CCGGACTCAACGTTGACATTTATCATAGTACGGTCCGTTAAACGTCGCTGCTCGAA 180
DB 713 CCGGATTCACCGTGGCATTTATCATAGTACGGTCCGTTAAACGTCGCTGCTCGAA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACCATTCACCGACGCGTTATGTTAACGGCGCAT 240
DB 773 AGCGATGCCGCTAAATCTGAAACGACCATTCACCGACGCGTTATGTTAACGGCGCAT 240
QY 241 GTAGGCGAGGTCGCGATTAATGTAATTAAGTCACTAGATGTTTCAAGAAATTAAT 300
DB 833 GTAGGCGAGGTCGCGATTAATGTAATTAAGTCACTAGATGTTTCAAGAAATTAAT 300
QY 833 GTAGGCGAGGTCGCGATTAATGTAATTAAGTCACTAGATGTTTCAAGAAATTAAT 892
DB 301 GCCACATCGACCACTGGAACGCTAAATCTCCATTTTACTGTCGCGCAATACGGCGGT 360
DB 893 GCCACATCGACCACTGGAACGCTAAATCTCCATTTTACTGTCGTCATACGGCGGT 952
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGGTAATGGTGGTCAAGTT 420
DB 953 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGGTAATGGTGGTCAAGTT 420
QY 421 GGTTCGCAACCAACGCGCGCTAATCACTATTA 455
DB 1013 GGTTCGCAACCAACGCGCGCTAATCACTATTA 1047

```

RESULT 8

LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
 DEFINITION Sequence 56 from patent US 5635617.
 ACCESSION 144908
 VERSION 144908.1 GI:2469621
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 UNCLASSIFIED.
 1 (bases 1 to 361)
 AUTHORS Doran, J. L., Kay, W. W., Collinson, S. Karen, and Clouthier, S. C.
 TITLE Methods and compositions comprising the agfa gene for detection of
 JOURNAL Salmonella
 Patent: US 5635617-A 56 03-JUN-1997;

FEATURES
 source Location/Qualifiers
 1..361
 /organism="unknown"
 /mol_type="unassigned DNA"

Query Match 62.3%; Score 284.2; DB 6; Length 361;
 Best Local Similarity 99.0%; Pred. No. 4.3e-75;
 Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 110 ATGCATCCGCGCCGACTCAACGTTGACGATTTATAGTAGCGTTCGGTAAACGTCGCGC 169
DB 47 ATATTCGCGCCCGACTCAACGTTGACGATTTATAGTAGCGTTCGGTAAACGTCGCGC 106
QY 170 TTGCTGCAACGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTATGTTA 229
DB 107 TTGCTGCAACGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTATGTTA 166
QY 230 ACGCGCGCATGTAAGCGCGGTCGCGATTAATGTAATGTAATGTAATGTAATGTAATGTA 289
DB 167 ACGCGCGCATGTAAGCGCGGTCGCGATTAATGTAATGTAATGTAATGTAATGTAATGTA 226
QY 290 TCAGAAATATATGCAACCATGACACGATGAAACGCTAAATCCGATTAATCTGCGCC 349
DB 227 TCAGAAATATATGCAACCATGACACGATGAAACGCTAAATCCGATTAATCTGCGCC 286
QY 350 AATACGCGCGTAAATACCGCGCGTGTATATCAAGCCGATCTGATTC 398
DB 287 AATACGCGCGTAAATACCGCGCGTGTATATCAAGCCGATCTGATTC 335

```

RESULT 9

CSP515700 2889 bp DNA linear BCT 24-JUN-2003
 LOCUS Citrobacter sp. fec2 csb gene, csd gene and csd gene.
 DEFINITION AJ515700.1 GI:31790491
 ACCESSION AJ515700.1
 VERSION AJ515700.1
 KEYWORDS csb gene; csb gene; csd gene; curlin-csb protein; nucleation component of curlin monomers; regulatory protein.

SOURCE Citrobacter sp. fec2
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.

REFERENCE 1
 Zogaj, X., Bokranz, W., Nintz, M., and Romling, U.
 Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract Infect. Immun. 72 (7), 4151-4158 (2003)
 2 (bases 1 to 2889)
 AUTHORS Romling, U.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-2002) Romling U., Microbiology and Tumorigenology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN

FEATURES

source

1..2889
 /organism="Citrobacter sp. fec2"
 /mol_type="genomic DNA"
 /isolate="fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:213763"
 /country="Germany"
 /complement(212..862)
 /gene="csbD"
 /complement(212..862)
 /gene="csdD"
 /complement(212..862)
 /gene="csbD"
 /function="necessary for transcription of the csbBA operon"

CDS

1
 /codon_start=1
 /transl_table=1
 /product="putative regulatory protein"
 /protein_id="CAD56670.1"
 /db_xref="GI:31790492"

```

/translation="MFNEFHSIHGHTLLITKPSLQATALLQHLKOSLALSGKLNIO
HSLDSSICIVLDVMEADKLIHYQDNLSRKNNIKTLLTPDPYRDIENMP
HNGVFPYATEDBQVRVSGLOGVAGECFQKSLASYLTHSGNRYNVSBSALLTHRE
KEILNKIRIGASNIENIARSLFISENTVKTHLYNLFKLIAYNRTOAVSWANDILRR"
1623..2078
/gene="csgb"
1623..2078
/gene="csgb"
/feature="nucleator"
/codon_start=1
/translation="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKKLLFMMFTLVGAPGIASTSYDLANSEYNAVNELSKSF
QAATIGOVGTNSAKROEGSKLSVVSQEGSNRAKVDGAINFAITAGSHSND
SISQSYGNTAMIIQKSGNKNANITGYOKTAVVQROSGMAIRVQR"
2119..2571
/gene="csga"
2119..2571
/gene="csga"
/feature="csga"
/function="curli subunit"
/codon_start=1
/translation="curlin-csga protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLIQVAFPAIIVSGSALAGVVPQWGGGGGGSSGPESTL
SIYSGVNNALAQSDARKSDITIHQGFNGNAGVGGSDNSTIDLTQNFKNATTI
DQWNGKSDIVTSQYGGHNAALVNOTASDSVLYHQVGFNNATANG"

ORIGIN
Query Match 61.1%; Score 278.4; DB 1; Length 2889;
Best Local Similarity 77.2%; Pred. No. 3e-73;
Matches 352; Conservative 0; Mismatches 101; Indels 3; Gaps 1;

QY 1 ATGAAACCTTTAAAGTGGAGCATTCGACGAATCGAGTTCGAGAGTCTTGCT 60
Db 2119 ATGAAACCTTTAAAGTGGAGCATTCGACGAATCGAGTTCGAGAGTCTTGCT 2178

QY 61 GGCCTATATGATCAAGCTGTTACCGGTGTTTACCCATGAATGCGACATCCGCG 120
Db 2179 GGCCTATATGATCAAGCTGTTACCGGTGTTTACCCATGAATGCGACATCCGCG 2235

QY 121 CCGGACATCAAGTGAAGATTATCAAGACGTTCCGTAACGCTGCGCTCTCGAA 180
Db 2236 CCGGACATCAAGTGAAGATTATCAAGACGTTCCGTAACGCTGCGCTCTCGAA 2295

QY 181 AGCGATGCGCGTAAATCTGAAGACCATTAACCAAGCGGTTATGTTACGGCGCGAT 240
Db 2296 AGCGATGCGCGTAAATCTGAAGACCATTAACCAAGCGGTTATGTTACGGCGCGAT 2355

QY 241 GTAGGCGAGGGTGGGATTAATAGTACTATTAAGTGAATGTTTTCAGAAATTAAT 300
Db 2356 GTAGGCGAGGGTGGGATTAATAGTACTATTAAGTGAATGTTTTCAGAAATTAAT 2415

QY 301 GCCACCATCGACCACTGGAAGCGTTAAACCTCCGATATTATGTCGCGCAATACGGCGCT 360
Db 2416 GCCACCATCGACCACTGGAAGCGTTAAACCTCCGATATTATGTCGCGCAATACGGCGCT 2475

QY 361 AATAAGCGCGGCTGGTATTATCAGACCGCATCTGATTCAGCGTAATGTTGTCAGAGTT 420
Db 2476 AATAAGCGCGGCTGGTATTATCAGACCGCATCTGATTCAGCGTAATGTTGTCAGAGTT 2535

QY 421 GGTTTGGCAACAAAGCGGCTTAACCAAGTATTAA 456
Db 2536 GGTTTGGCAACAAAGCGGCTTAACCAAGTATTAA 2571

RESULT 10
CFR515701 2920 bp DNA linear BCT 24-JUN-2003
LOCUS CFR515701
DEFINITION Citrobacter freundii csgB gene, csgA gene and csgD gene.

```

```

ACCESSION AJ515701
VERSION AJ515701.1 GI:31790495
KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE Citrobacter freundii
ORGANISM Citrobacter freundii
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
REFERENCE 1
AUTHORS Zogaj,X., Bokranz,W., Nimitz,W. and Romling,U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
JOURNALS Infect. Immun. 72 (7), 4151-4158 (2003)
2 (bases 1 to 2920)
Romling,U.
DIRECT SUBMISSION
SUBMITTED (12-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
SOURCE
1..2920
/organism="Citrobacter freundii"
/mol_type="genomic DNA"
/isolate="fec4"
/isolation_source="faeces"
/species_host="Homo sapiens"
/db_xref="taxon:546"
/country="Germany"
/gene="csgD"
/complement(211..861)
/gene="csgD"
/function="necessary for transcription of the csgBA operon"
/codon_start=1
/translation="putative regulatory protein"
/protein_id="CAD56673.1"
/db_xref="GI:31790496"
/translation="MFNEFHSIHGHTLLITKPSLQATALLQHLKOSLALSGKLNIO
RSLDSSICIVLDVMEADKLIHYQDNLSRKNNIKTLLTPDPYRDIENMP
HNGVFPYATEDBQVRVSGLOGVAGECFQKSLASYLTHSGNRYNVSBSALLTHRE
KEILNKIRIGASNIENIARSLFISENTVKTHLYNLFKLIAYNRTOAVSWANDILRR"
1624..2082
/gene="csgb"
1624..2082
/gene="csgb"
/function="nucleator"
/codon_start=1
/translation="nucleation component of curlin monomers"
/protein_id="CAD56674.1"
/db_xref="GI:31790497"
/translation="MKKLLFMMFTLVGAPGIASTSYDLANSEYNAVNELSKSF
QAATIGOVGTNSAKROEGSKLSVVSQEGSNRAKVDGAINFAITAGSHSND
SISQSYGNTAMIIQKSGNKNANITGYOKTAVVQROSGMAIRVQR"
2123..2572
/gene="csga"
2123..2572
/gene="csga"
/function="curli subunit"
/codon_start=1
/translation="curlin-csgA protein"
/protein_id="CAD56675.1"
/db_xref="GI:31790498"
/translation="MKLIQVAFPAIIVSGSALAGVVPQWGGGGGGSSGPESTL
SIYSGVNNALAQSDARKSDITIHQGFNGNAGVGGSDNSTIDLTQNFKNATTI
DQWNGKSDIVTSQYGGHNAALVNOTASDSVLYHQVGFNNATANG"

ORIGIN
Query Match 50.3%; Score 229.2; DB 1; Length 2920;
Best Local Similarity 70.6%; Pred. No. 2.8e-58;
Matches 322; Conservative 0; Mismatches 128; Indels 6; Gaps 1;

```

QY 1 ATGAACCTTTTAAAGTGGACATTCGACCAATCGTAGTCTTGGCAAGTCTTGGCT 60
 Db 2123 ATGAACCTTTTAAAGTGGACATTCGACCAATCGTAGTCTTGGCAAGTCTTGGCT 2182
 QY 61 GCGCTATGATCAGCTGTTTACCGCTTCTTACCATGAAATGACATGCGATCGCGC 120
 Db 2183 GCGCTATGATCAGCTGTTTACCGCTTCTTACCATGAAATGACATGCGATCGCGC 120
 QY 121 CCGACATCAGCTTACGATTTATCAGTACGCTTCCGCTACGCTGCGCTTCTTGGCA 2236
 Db 2237 CCGACATCAGCTTACGATTTATCAGTACGCTTCCGCTACGCTGCGCTTCTTGGCA 180
 QY 181 AGCAGTCCGCTTAAATCTGAAGACCATTAACCAAGCGTTATGATGATGATGATGAT 2296
 Db 2297 AGTATGCGCTTAAATCTGAAGACCATTAACCAAGCGTTATGATGATGATGATGAT 240
 QY 241 GTAGCCAGGCTGCGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2356
 Db 2357 GTAGCCAGGCTGCGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 301 GCCACATGACGACGATGAAACGCTAAATCTCCGATTTACTGTCGCGCAATGCGCGCT 2416
 Db 2417 GCCACATGACGATGAAACGCTAAATCTCCGATTTACTGTCGCGCAATGCGCGCT 360
 QY 361 AATAACGCGCGCTGCTTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 2476
 Db 2477 CCGACATGACGATGAAACGCTAAATCTCCGATTTACTGTCGCGCAATGCGCGCT 420
 QY 421 GGTGTTTGGCAACGACGCGCTAACCATGATTTAA 456
 Db 2537 GGTGTTTGGCAACGACGCGCTAACCATGATTTAA 2572

RESULT 11
 LOCUS AF275733 1711 bp DNA linear BCT 15-MAY-2001
 DEFINITION Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
 ACCESSION AF275733
 VERSION AF275733.1 GI:14039399
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 QY 1 (bases 1 to 1711)
 Db Mutations in the csgD promoter associated with variations in curl
 QY Expression in certain strains of Escherichia coli O157:H7
 Db Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
 QY 11319125
 Db 2 (bases 1 to 1711)
 QY Uhlrich, G.A., Keen, J.E. and Elder, R.O.
 Db Submitted Direct Submission
 QY Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
 Db Research Center, State Spur 18D, Clay Center, NE 68933, USA
 FEATURES
 source
 1. 1711
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="43895 Red Variant"
 /serotype="O157:H7"
 /db_xref="taxon:562"
 /gene="csgB"
 /gene="csgB"
 /gene="csgB"
 /transl_table=1
 /codon_start=1
 /product="CsgB protein"
 /protein_id="AAK53211.1"

ORIGIN
 Query Match 49.4%; Score 225.4; DB 1; Length 1711;
 Best local similarity 69.7%; Pred No. 3,8e-57;
 Matches 320; Conservative 0; Mismatches 136; Indels 3; Gaps 1;
 /db_xref="GI:14039400"
 /translation="MKMLFPMWLTILGADGIAAAGYDLANSYFAVNELSKSSFN
 QALIGAGNTNNSQILROGSKSLAIVADGSSNRKRIQOTGYDNLAYIDQAGSANDA
 STISQAVGNTAMIIQKSGSNKANTYGTOKAIVVQROGMAIRVTOG"
 1253.1711
 /gene="csgB"
 1253.1711
 /gene="csgB"
 /gene="csgB"
 /codon_start=1
 /transl_table=1
 /product="CsgB protein"
 /protein_id="AAK53212.1"
 /db_xref="GI:14039401"
 /translation="MKLLVFAIAIIVSGSLAGVVPYGGGNGGNGGNGGNSPNS
 ELNITVOYGNSNLAIAQADARNSDLTIQHGNGAGNDVGGSSDSSIDLTOGFGNSA
 TLIDMNGKSHMTVQFQGGNGAIVDQASNSVTNVTVGFGNATAHQY"

QY 1 ATGAACCTTTTAAAGTGGACATTCGACCAATCGTAGTCTTGGCAAGTCTTGGCT 60
 Db 1253 ATGAACCTTTTAAAGTGGACATTCGACCAATCGTAGTCTTGGCAAGTCTTGGCT 60
 QY 61 GCGCTATGATCAGCTGTTTACCGCTTCTTACCATGAAATGACATGCGATCGCGC 1312
 Db 1313 GCGCTATGATCAGCTGTTTACCGCTTCTTACCATGAAATGACATGCGATCGCGC 117
 QY 118 GCGCCGACATCAAGTGGAGATTTATCACTAGTACGCTTCCGCTAACGCTTGGCTG 1372
 Db 1373 GCGCCGACATCAAGTGGAGATTTATCACTAGTACGCTTCCGCTAACGCTTGGCTG 177
 QY 178 CAAAGCATGCGCGTAAATCTGAAGACCATTAACCAAGCGCTTATGATGATGATGAT 1432
 Db 1433 CAAAGCATGCGCGTAAATCTGAAGACCATTAACCAAGCGCTTATGATGATGATGAT 237
 QY 238 GATGAGCCAGGCTGCGGATTAATGATGATGATGATGATGATGATGATGATGATGATG 1492
 Db 1493 GATGAGCCAGGCTGCGGATTAATGATGATGATGATGATGATGATGATGATGATGATG 297
 QY 298 AATGCCACATGACCAATGTAAGCAATTAACCAAGCGCTTATGATGATGATGATGAT 1552
 Db 1553 AATGCCACATGACCAATGTAAGCAATTAACCAAGCGCTTATGATGATGATGATGAT 357
 QY 358 GGTATTAACGCGCGCTGCTTAAATCAAGACCGCATCTGATTCAGCGTAAATGCTCAG 1612
 Db 1613 GGTATTAACGCGCGCTGCTTAAATCAAGACCGCATCTGATTCAGCGTAAATGCTCAG 417
 QY 418 GTTGTGTTTGGCAACGACGCGCTAACCATGATTTAA 456
 Db 1673 GTTGTGTTTGGCAACGACGCGCTAACCATGATTTAA 1711

RESULT 12
 LOCUS AE005315 10190 bp DNA linear BCT 21-MAR-2001
 DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
 ACCESSION AE005315
 VERSION AE005315.1
 KEYWORDS
 SOURCE
 ORGANISM Escherichia coli O157:H7 EDL933
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 QY 1 (bases 1 to 10190)
 Db Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
 Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
 Grodzicki, E.J., Davis, N.W., Lim, A., Diallanta, E., Potamousis, K.,
 Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,


```
gene /db_xref="taxon:83334"
150..2258
/gene="ECs1032"
150..2258
/note="similar to YCBY ECOLI g1|1787181 percent identity
99 in 702 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="putative oxidoreductase"
/protein_id="BAB34455.1"
/db_xref="GI:13360492"
/translation="WNSLPASTARGLEELKTELENI.GAVCOVVOGSHRKQDRLV
YQSIMSLASRIMLPGECKYSLDLIVGQAINWTEMRPQATRVHSGNDITV
RNSOTGAKVDAIVDAITRNKLPENVDADPDIRVNVMLKETAIALDISGGLH
LRGRDRAGIAPIKETTLAAIYMRGQWPGTPLDPMCGSTLLI.EAAMI.LTDRAPIG
LRGHWGFSGAHQHDEAIQWEVYAEQTPARKGLAESSHFYGSDDARVIGRATNAR
LAGIGELITFEVKVAOLTNPLPEPYGTVLNPPYGBRLSEPALIALHSLGRIMK
NPGSGWNLSPASPDILSCIOLRADKQYKAKNGPLDCVOKNHYAESTPDSKPMVA
EDYANRLAKNLUKPEKAROSIGECYRYADLPREYNAVDRYADWYVGYEAPPKTI
DAHKRQRKFDITIAFTISVLGIAPKLVLTREKQKNQYOKLGEKEFLEVEYNA
HLWNLTDYLDLDTGLFLDRIARMLQMSKQKDFLNLSTYSATVAHAGLGARSTTT
VDMSTYLEMAERNLRLNGLTGRARHLIQADCLAMLRANEQFDLIFIDPTFSKSR
MEDAFDVOHDIALMKDKRLLRAGGTIMFSNNKRGFRMDLDGLAKGLKAKQETQKT
LSODPARROIHNCMLTAA"
2270..4177
/gene="ECs1033"
2270..4177
/note="similar to UUP ECOLI g1|1787182 percent identity 99
in 635 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="putative ATP-binding component of a transport
system"
/protein_id="BAB34456.1"
/db_xref="GI:13360493"
/translation="WSLISWHGAMLSFSDAPLNDNAELIHEDNERVCIVGNAGKST
LMKILNREGQIDGDIIEYODLIVARLOQDPRNVEGVDVFAEGEBOEAYLRHY
DISRLVNDPSEKINELAKYQEQLDHNLINQLNRLINEVLAQLDIPNVASSISGG
WLRKALGRALVSNPRVLLDDEPTNDLLEITIDMEGLKTFNGTIIISIDRSIRA
MATRIIVLDRCGLVTPGNYOVLLEKEALRVELONAFERDRLAOEYVIRGICIA
RRTREBGRVRLKAMRBERGEREMGTAKQVEBASGKIVEMEDVCYQVQKOL
VKDPSAQLVMDGKIALIGPNCGKTTLLKLMGOLQADSGRIHVGTLEVAYFDQHRA
ELDDPKTVMDNLAGKOEQVWNGKRRHVLGYLOFLFPPKRAMTPVRLASGEBNRL
LARFLKPSNLLIIDEPTNDLDEVELLEELIDSYOGTVLVSHDRQFNVNTYECW
IFEGGKIGRYVGVGHARQGOEOYVALKOPAVKKNERPAPKATVYRSSSKLYKL
ORELEBOLLEDELEAKLEALQTOYVADASFSGQHEQYQKVLADMAAAEQLEQAFER
WEYLEALKNGG"
4307..5560
/gene="ECs1034"
4307..5560
/note="similar to PQIA ECOLI g1|1787183 percent identity
99 in 417 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="paraquat-inducible protein A"
/protein_id="BAB34457.1"
/db_xref="GI:13360494"
/translation="WCBHHAAKHILSCQCMVALPRLHQAQACPRCCTTLTVAM
DAPRQPTAVLAALFMLLSNLPFPVMMNVAQVSEITLLEIGVLFSEYASIVSF
FLFVQVLPAPALITLILVNRABLPVLAKEQLARVAFQETLWMAELFLGASVGF
KLMAVSGISGVSSFLPCLFCVLOLRACQCDRMLMDLIPMLSELPLKRGVTSR
GLRSGSCCTAIIIPADEVPCRCGTCGKVRBNLSLMDTALIVSIMYLPANLIPM
VTDLIGSMPESTIIAGVILLMSBGSYPANVITFLASINVPFLKAIATLMLCWDKGS
KRDERKHLITVEVEFVGKMSIDVFVAIVSLAVRKGKLSITYPAMGALPFAIVIN
TFPSAMTFDPLSLWDROPESEHES"
5565..7205
/gene="ECs1035"
5565..7205
```

```
gene 5565..7205
CDS /gene="ECs1035"
/note="similar to PQIB ECOLI g1|1787184 percent identity
99 in 546 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="paraquat-inducible protein B"
/protein_id="BAB34458.1"
/db_xref="GI:13360495"
/translation="WESNNGEAKIOKYNMSPWVIFPIVTLIGAWLVFYHSHQPS
VTLLTANAEIGEGKTIKSRSDVGVVGSATLADDLTHVEIKARLNGMKRLAKOT
VFWVVKPOEGREGISGLTISGVYIELOPAKSGSKMDKDLDSPLAPADAKITR
VLDSKXAGLSRPGPVPRGTRVSVSTSTDTQKRNISYQLFINAPDRLVTSVRF
WDSGIAVLDLTSAMRVEMSLITLISGVSFDVBEGLDQVAPAPKAPVLYDDQS
IODELYVDHIDYLMFEDSVAGLOPQAPVEFRGRLGTVSVPPFAPMWRQTFNDYR
IPVIRIIEPERLKKQLGEMNDVYGLSELKRGGLRGLKTNLVTGALYVLDDEFNT
PATIGIEFNFOYQIIPVSGLAQIQORLMALDKINKLPLNPMIEQATSTLSORT
MKNLOTTLDNKNKILASOSMOQLPTDMOSTIRELNRMGQFQPSAAVNMVADMORL
DOVIREQPVYKTLNESNMLVFBAKDKDPEPRAKQ"
7217..7765
/gene="ECs1036"
7217..7765
/note="similar to YMBR ECOLI g1|1787185 percent identity
98 in 182 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB34459.1"
/db_xref="GI:13360496"
/translation="WITIAALMLAGSGSGEINKRYQLPVQSGTQSTRSQGNRLMYE
QVAVPDYLAGGVVYQTSYDVKYVANNLWASPLDQDLRLTLVNLSTQLGQWVASQ
PLGSAQDTLVNTVTFEFGRYDGKIVYGVGMLNMQOLIKRPFLEGVQTDGYDEM
KVLGVNSQEMAASTAGRIKRLP"
8021..8188
/gene="ECs1037"
8021..8188
/note="identical to RMP ECOLI g1|1787186 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1
/product="ribosome modulation factor"
/protein_id="BAB34460.1"
/db_xref="GI:13360497"
/translation="MKRQKRDLERAHQRGVQAGIAGRSKEMCPYQTLNRSQWLGM
REMAADRVVMA"
8776
/gene="ECs1038"
8776
/complement (8258..8776)
/gene="ECs1038"
8776
/note="similar to PABA ECOLI g1|1787187 percent identity
100 in 172 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=1

Query Match 49.4%; Score 225.4; DB 1; Length 327773;
Best local similarity 69.7%; Pred. No. 6e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 3; Gaps 1;

QY 1 ATGAACCTTTAAAGTGGAGCATTCGATGCTTTCGACAGTCTCGCT 60
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 326421 ATGAACCTTTAAAGTGGAGCATTCGATGCTTTCGACAGTCTCGCT 60
| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 GCGCTATATGATCAGCTGTTTACCCGCTGTTTATCCCA---TGAATGGCAGCATGATCC 117
| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 326481 GGTGTGTTCTCTCAGTACGGGCGGCGTGGCGGTACCAACGGTGTGGCGGTATTAACAGC 326540
| | | | | | | | | | | | | | | | | | | | | | | | | |
QY 118 GCGCCGACATCAAGTTGAGCATTTACGATGCGTTCGCGTACGCTGCGCTGCTG 177
| | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

/codon_start=1
/transl_table=11
/product="assembly /transport component in curl
production"
/protein_id="CA62277.1"
/db_xref="GI:1147559"
/db_xref="NCBI:P52103"
/db_xref="SWISS-PROT:P52103"
/translation="MORFLVAVNLGGCLTAPKPEARPTLMPAQSVDLTHLP
PTGTLFISVYNIODETGFQFPYPAANSTLTPVPSATRAMLVTKDSRNPFLIEQGLC
NLNLRKRIIRAQOENAVLNNRIPLDSLTAINWEGSIIGESNVSGSGVAGRYRG
IGADTVOYDODAVNLRYVNSTEILISVNTSKTILSYEVAGVRFIDYQRLLEESE
VGTSTNPEMLCLMAITGTIVFLINDIGRLMDLQNAERQNDILVKYHNMSVPE
S"
complement(993..1409)
/gene="csf"
/gene="csf"
complement(993..1409)

```

production"
/protein_id="CAA62278.1"
/db_xref="GI:1147560"
/db_xref="SWISS-PROT:P52104"
/translation="MRVKAIVVLMILISPLSMACTYTFQFRPNRFGCINPNNAPFLINL
AQAQNSKSDSYNDFCIGTSPALNPFQAQIQSITLGLGLLINTIGKPFQMTVNTYIT
DIARDQQLQNVYDQKGTGOSTTIVSGLDNNSTD"
complement(1434..1823)
/gene="cse92"
complement(1434..1823)
CDS

```

/transl_table=11
/product="assembly /transport component in curl
production"
/protein_id="CAA62279.1"
/db_xref="GI:1147561"
/translation="MKRLRWIVAAEPLFPAAGNLHAVEVEVGLITDHTVSIGIDHPY
RAVSDWEDLYDGNLTINERPSAMGSMITTYNDQVIFOTPLPKKDFEKTVPALL
IQEELNLRROIALLSTGDLHAIDPF"
complement(1828..2478)
/gene="csap"
/complement(1978..2478)

```

```

/codon_start=1
/strand1_table=11
/product="putative regulatory protein"
/protein_id="CMA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translation="MFNEVSHIGHLTLLTSLKSLQATLTLQSLATLTKLNLNQ
RSLIDISSGSIILLDMEDADKKLIHWQDPTLSKSNKNNITLILNTPEDYPIEDNNE
HINFGYEMEDQERVNGVLCQVARGECYFQKASLALITHSGNYSTESLALTHRE
KEINIKRIGASNNNEIKRSTFIENETVKTLNLFKLAKVAKRQTVASMANDNLRR"
/complement(12626)
/misc_feature
    complement(12626)
    note="transcriptional start site for csdBFG operon"
    3140
/misc_feature
    note="transcriptional start site for csGBA operon"
    3233..3688
/gene="csGB"
/gene="csGB"
    3233..3688
CDS

```

```

/translation="MKKTLFMMLTILGAPGIAAAGYDLANSEYNFAVNLSSKSPN
QAATIGAGCTNNASQLROGSKLAVVADEGSSNPAKIDQGDVYTLATIDAGSANDA
SISGAYCNTAMTIIOKSGGNKAMITQYGIKTAIVQKQSMALNVTOR"
3729..4184
/gene="csga"
3729..4184
/feature="csga"
/feature="major subunit"
/codon_start=1
/translation="curliin/csga protein"
/product="curliin/csga protein"
/protein_id="CAA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="MKLKVAAIAIAIVSGSALAGVPOYGGGNGHGGGNNSGPNSL
LNIYOYGGNSALAQTDARNSDLTIQHGNGNDVGGGSDSDS1DITQGFQNSAT
LDOWNGNSEMTVQFGGGNGAAVDQTAANSNVVTQVGFQNNATAHQY"
3729..3788
/gene="csga"
4243..4575
/gene="orfC"
4243..4575
/gene="orfC"
/codon_start=1
/translation="putative curli production protein"
/product="putative curli production protein"
/protein_id="CAA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="MNTLLLAALSSQITFTNTQGGDYTTIPEVTLTQSLCVOIL
SLRSGSGSQTQKQKTLISLPANQPIALTKLSLNI SPDRKVIYVTVSDGSLHLSQO
WPSPSKS"

```

ORIGIN

Query Match 49.1%; Score 224; DB 1; Length 4680;
 Best Local Similarity 68.2%; Pred. No. 1.1e-56;
 Matches 311; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGACGATTCGACGCAATCGTAGTTCTGCGAGTCTTGCT 60
DB 3729 ATGAACTTTTAAAGTGGACGATTCGACGCAATCGTAGTTCTGCGAGTCTTGCA 3788
QY 61 GGCGCTATGATCACTGCTGTACCCGTTGTGTACCCATGAAATGACATGATCCGCC 120
DB 3789 GGCTGTGCTCTCACTGACGCGCGCGGTAAACACGCTGTGCGGTAAATAGCCGC 3848
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAAGCTGCTGCTGCA 180
DB 3849 CCAAACTTCTGAGTGAACATTTACAGTACGCTGCGGTAACTCTGCACTTCTCTGCA 3908
QY 181 AGCGATGCCGCTAAATCTGAAACGACATTCACCGAGCGGTTATGTTAACGGCGCAT 240
DB 3909 ACTGATGCCGCTAACTGACTGATCTATTAACCGAGCATGCGCGGTTAATGTCAGAT 3968
QY 241 GTAGCCAGGCTGCGGATTAATAGTACTTGAACCTGACTGAAATGTTTCAAAATAT 300
DB 3969 GTTGTACGGGCTCAGATGACGCTCAATCACTGACCAACGCGTGTGGTAAACGC 4028
QY 301 GCCACATCGACGAGTGAACGCTAAACTCCGATATTATCTGTCGCGCAATAGCGCGT 360
DB 4029 GCTACTCTGATCACTGGAACGCGCAAAATCTGAAATGACGCTTAAACATTTGCTGCT 4088
QY 361 AATAACGCCGCGCTGTTAATCAACCGCATCTGATTCACGCTAATGTCGCTGAGTT 420
DB 4089 GCGAATCGGTGCTGAGTGAACGACATCTGATTAACCTCTCCGTCACGTGACTCAGGT 4148
QY 421 GCTTTGGCAACAACGCGCAACGCTAACAGATTTAA 456
DB 4149 GCGTTGTAAACAACGCGCAACGCTAACAGATTTAA 4184

```

RESULT 15

```

AE000205
LOCUS AE000205 10346 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 95 of 400 of the complete
genome.
ACCESSION AE000205 U00096
VERSION AE000205.1 GI:1787265
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 10346)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
97426617
9278503
2 (bases 1 to 10346)
Blattner, F.R.
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 10346)
Blattner, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 10346)
Plunkett, G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Fredrick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using Genemark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.
FEATURES
source
1..10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="MG1655"
/db_xref="taxon:83333"

```

```

repeat_region complement(<1. .48)
promoter /note="IS3" .
promoter /note="factor Sigma70; predicted +1 start at 1094689"
gene /note="factor Sigma70; predicted +1 start at 1094716"
CDS /gene="b1028"
/gene="b1028"
/function="orf; Unknown"
/note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
to 58 residues of an approx. 188 aa protein RAS2_CANAL
SW: P43098"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/db_xref="GI:178726"
/translation="MEDGVQAMRDYLAGLDIASPEHOVLNNVTAKSEVAPSTIKENLS
LHLHTVKTESLDTFLNMPVPAFLISNKPYLGNMLNDPAGVDQQRVHCKKAFSD
AKVFX"
275. .299
/gene="b1028"
/note="factor Sigma32; predicted +1 start at 1094983"
389. .1375
/gene="ycdu"
/note="synonym: b1029"
389. .1375
/gene="ycdu"
/function="orf; Unknown"
/note="o328; This 328 aa ORF is 19 pct identical (4 gaps)
to 191 residues of an approx. 288 aa protein Y320_MYCE
SW: P47562"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"
/translation="MIPYLPIRPRQDRNLIYYAIGLIGFYWNAQFPSEDI
LIMLSMWGLVYVNFIDLKAYAVKCVKNIDFSWPKKQNHKLEFLQPLVAGTIS
SLTRVYLSVCISTVYVNLITISPLNSDSFVTEGOMLPKSIITALLICGVATNLF
PIRPSKRIAFGLRLOBIIDLFESSENLSTFPKPLMLRFLITLVYEWMTLVSVL
ALVLEKRIAFEEYFLLCYVPCLIYFYFCRPLMNDPMACDMYFRKGFHNK"
1253. .1280
/gene="ycdu"
/note="central position to predicted promoter: -104"
1253. .1280
/bound_moiety="Deor predicted site"
/gene="ycdu"
/note="No predicted promoter"
/bound_moiety="Deor predicted site"
1337. .1365
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1096049"
1494. .1745
/gene="b1030"
/gene="b1030"
/function="orf; Unknown"
/note="o83; This 83 aa ORF is 25 pct identical (7 gaps)
to 81 residues of an approx. 616 aa protein ALBU_RABIT
SW: P49065"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGRGVYTKKKKPUYSYELFRYGGEGIDSLRSPCGPAHC
VHGLNMLSPVVDPRSGLLIPVCAIHEKKAFTVRL"
1513. .1926
/gene="ycdu"

```

```

CDS
/note="synonym: b1031"
1513. .1926
/gene="ycdu"
/function="putative structure; Not classified"
/note="o137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein RS37_YEAST
SW: P05759"
/codon_start=1
/transl_table=1
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKSPYRFTSSSLNMAVRGGLTPRPPAGSLTATFYCPT
GKRLSTPGKGSFSPRCVOYTKKPYLSYELFPPKGGGIDSLRSPCGPAHCVHGL
SNMLSPVBRSGLLIPVCAIYEKKARFVRL"
complement(2111. .2198)
/gene="serX"
/note="synonym: b1032"
complement(2111. .2198)
/gene="serX"
/product="tRNA-Ser"
/note="anticodon: GCA; CG Site No. 17878"
2269. .2296
/function="tRNA; tRNA"
/note="factor Sigma70; predicted +1 start at 1096980"
2338. .2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355. .2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393. .3370
/gene="ycdu"
/note="synonym: b1033"
2393. .3370
/gene="ycdu"
/function="putative enzyme; Not classified"
/note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPBA_CORGL
SW: P45637"
/codon_start=1
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="KRIKFAQFSGVSDIIFYHPTPTOMWIBLARKAIPOARVAM
KSGDNDADYALVWHPPEVEMLAGDILKAVFALGAGVDILSKLAHPMLNPSPLR
LEDTMGEGMOEYAVSOVLHMERFDDYRIOONSHMOPDEVREDPTTIGTAGVL
GSKVAQSLQTMRFPLRCWSRTERKSWPGQSGRERUSATLSQCRVLINLEPTEPV
GILNQLLEKLPDGAIVLNLARGVHVVDLILALDGRKAGALVDVFNPEPLPEPSP
LMQHPVITTPHVAITRPAVAVEISTITQLLEKGERVCGQVDRARGY"
3278. .3308
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1097992"
3383. .3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396. .3423
promoter
promoter
Query Match 49.1%; Score 224; DB 1; Length 10346;
Best Local Similarity 68.2%; Pred. No. 1.2e-56;
Matches 311; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGAGCGACATTCGACGACATCGTAGTTCGCGAGCTCGT 60
DB 8993 ATGAACCTTTAAAGAGCGACATTCGACGACATCGTAGTTCGCGAGCTCGT 60
QY 61 GAGCTATGATCGAGCGGTACCCGCTGTTGTTACCATGAAATGCGACATCGCGC 120
DB 9053 GATGTTGTTCTCGATACGCGCGCGGGAACCAAGGTGCGGATTAATAGCGCG 9112
QY 121 CCGAGCTCAAGTTGAGCATTTATCAGTACGCTCGCGTAAAGCGCGCTGCTGCA 180
DB 9113 CCAATTCGAGCGAACAATTACAGTACGCTGCGGATCTGCACTGCTGCA 9172
QY 181 AGCGATGCCGTAATCTGAACGACATTACCCAGAGCGGTTATGTAAAGCGCGCAT 240

```

```

Db | 9173 | ACTGATGCCCGTAACTGACTTGACTATTACCCAGCATGGCGCGGTAAATGTCAGAT 9232
Qy | 241 | GTAGGCCAGGCGCGGATTAATGTAATACTGACTGAGATGTTTCAGAAATTAAT 300
Db | 9233 | GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCAAGTGCGCTCGGTAAACAGC 9292
Qy | 301 | GCCACCATCGACCACTGGAACGCTAAAACTCCGATATTACTGTCGCGCAATACGGCGGT 360
Db | 9293 | GCTACTCTTGATCAGTGAAACGGCAAAATTTCTGAAATGACGTTAAACAGTTGCGTGGT 9352
Qy | 361 | AATAACGCCCGCGCTGTTAATCAGACCGCATCTGAATTCAGCGTAATGATGCGTCAGGTT 420
Db | 9353 | GGCACAGGCTGCTGACGTTGACAGACTGATCTTAATCTCTCCGTCACAGTGAATCAGGTT 9412
Qy | 421 | GCTTTGGCAACAGCCAGCGCTAACCAATTTAA 456
Db | 9413 | GGCCTTGGTAAACGACGACCGCTCATCAGTAACTAA 9448

```

Search completed: March 15, 2004, 22:50:09
 Job time : 1965.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456
Sequence: 1 atgaactcttaaaagtcgc.....ccacgcctaccagtattaa 456.Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001s:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	3	AAC64624 Agfa::PT3
2	403.2	88.4	456	2	AAQ87467 Agfa sequ
3	403.2	88.4	456	2	AAAT74142 Salmonell
4	403.2	88.4	456	3	AAAC64617 Salmonell
5	360	78.9	456	3	AAAC64626 Agfa::PT3
6	353.6	77.5	456	3	AAAC64625 Agfa::PT3
7	352	77.2	456	3	AAAC64628 Agfa::PT3
8	352	77.2	456	3	AAAC64622 Agfa::PT3
9	352	77.2	456	3	AAAC64629 Agfa::PT3
10	352	77.2	456	3	AAAC64623 Agfa::PT3
11	348.8	76.5	456	3	AAAC64630 Agfa::PT3
12	348.8	76.5	456	3	AAAC64627 Agfa::PT3
13	345.6	75.8	456	3	AAAC64631 Agfa::PT3
14	284.2	62.3	361	2	AAQ73066 Agfa sequ
15	284.2	62.3	361	2	AAAT74141 Salmonell
16	222.4	48.8	456	3	AAAC64619 Escherich
17	222.4	48.8	456	9	AAAC64613 Agfa::PT3
18	210	46.1	646	2	AAAC64617 Agfa::PT3
19	149.6	32.8	359	2	AAAC64617 Agfa::PT3
20	51.2	11.2	100	7	AAAC64617 Agfa::PT3
21	49	10.7	78	9	AAAC64617 Agfa::PT3
22	48.2	10.6	78	9	AAAC64617 Agfa::PT3
23	48	10.5	48	3	AAAC64621 Agfa::PT3

24	48	10.5	78	3	AAAC64610 Agfa (SEF
25	43.2	9.5	48	3	AAAC64615 S. enteri
26	42.4	9.3	100	7	AAAC64615 S. enteri
27	41.6	9.1	72	9	AAAC64615 S. enteri
28	41.2	9.0	78	3	AAAC64605 Sefa (SEF
29	40.8	8.9	100	7	AAAC64605 Sefa (SEF
30	40.6	8.9	78	3	AAAC64606 Sefa (SEF
31	35.6	7.8	1965	4	AAAC64606 Sefa (SEF
32	35.6	7.8	1965	7	AAAC64606 Sefa (SEF
33	35.4	7.8	3411	5	AAAC64606 Sefa (SEF
34	35.4	7.8	3412	5	AAAC64606 Sefa (SEF
35	34.6	7.6	456	3	AAAC64620 Agfa::PT3
36	34.6	7.6	1506	6	AAAC64620 Agfa::PT3
37	34	7.5	456	3	AAAC64618 Agfa::PT3
38	34	7.5	2000	7	AAAC64618 Agfa::PT3
39	34	7.5	2583	7	AAAC64618 Agfa::PT3
40	34	7.5	2886	7	AAAC64618 Agfa::PT3
41	34	7.5	3300	3	AAAC64618 Agfa::PT3
42	33.6	7.4	549	3	AAAC64618 Agfa::PT3
43	33.4	7.3	3990	2	AAAC64618 Agfa::PT3
44	33.4	7.3	5100	2	AAAC64618 Agfa::PT3
45	33.4	7.3	7766	2	AAAC64618 Agfa::PT3

ALIGNMENTS

RESULT 1
AAC64624 standard; DNA; 456 BP.
ID AAC64624;

26-FEB-2001 (first entry)

Agfa::PT3#3 DNA sequence SEQ ID NO:15.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
vaccine; immune response; immunogen; ds.

Salmonella enteritidis.

Escherichia coli.

Synthetic.

WO200060102-A2.

12-OCT-2000.

05-APR-2000; 2000WO-CA000356.

05-APR-1999; 99US-0127888P.

(UVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay WW;

WPI; 2000-672631/65.

P-PSDB; AAB36348.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence

which encodes foreign epitope or antigen, expresses recombinant Agfa

protein useful for eliciting immune response in animal.

Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
segment of the gene has been replaced by a segment of a foreign DNA
sequence which encodes a foreign epitope or antigen. Also described are:
(1) use of thin aggregative fimbriae (SEF1/TFP) nucleation depended
assembly system of strains of Salmonella, Escherichia coli and
Enterobacteriaceae for the production of fimbriae comprising recombinant
Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
directing recombination of a recombinant gene into the chromosome of the

homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrin subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrin are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.5e-141;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATGCTGTTCTGGAGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATGCTGTTCTGGAGCTGCTGCT 60
QY 61 GGGCTCTATGATGAGCTGGTTACCGCTGTTTACCCATGAAATGGCAATGCATCCGCG 120
DB 61 GGGCTCTATGATGAGCTGGTTACCGCTGTTTACCCATGAAATGGCAATGCATCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTCCGCTACGCTGGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTCCGCTACGCTGGCTGCTGCA 180
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCGAT 240
DB 181 AGCGATGCGCGTAAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCGAT 240
QY 241 GTAGGCCAGGCTGCGATTAATAGTACTTTGAACTGAGTCAATGTTTCAAGAAATAT 300
DB 241 GTAGGCCAGGCTGCGATTAATAGTACTTTGAACTGAGTCAATGTTTCAAGAAATAT 300
QY 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
DB 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAAATGTTGCGTCAAGTT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAAATGTTGCGTCAAGTT 420
QY 421 GGTTTGGCAACACGCCACGCTTAACCAAGTATTTAA 456
DB 421 GGTTTGGCAACACGCCACGCTTAACCAAGTATTTAA 456

```

RESULT 2

AA087467 ID AA087467 standard; DNA, 456 BP.

XX AA087467;

DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)

DE Agfa sequence.

XX *Salmonella*; Agfa; vaccine; genetic immunization; ds.

XX *Salmonella*.

Key Location/Qualifiers
CDS 1..454
/tag=a
/note="Agfa"

MO9425598-A2.

10-NOV-1994.

26-APR-1994; 94MO-IB000207.

26-APR-1993; 93US-00054452.

(UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.

(KING/) KING J.

Kay WW, Collinson SK, Clouthier SC, Doran JL,

WPI; 1994-358275/44.

P-PSDB; AAR74625.

Eliciting an immune response to *Salmonella* - using attenuated *Salmonella* strains, vector constructs, or comps. confg. fibrin type proteins.

Disclosure; Fig 7B; 95pp; English.

The DNA encodes the *Salmonella* Agfa protein. The DNA and isolated proteins are used in genetic immunization and vaccine compositions, respectively, to elicit an immune response to *Salmonella* in animals (e.g. food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN field.)

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.4%; Score 403.2; DB 2; Length 456;
Best Local Similarity 92.8%; Pred. No. 2.9e-123;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATGCTGTTCTGGAGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATGCTGTTCTGGAGCTGCTGCT 60
QY 61 GGGCTCTATGATGAGCTGGTTACCGCTGTTTACCCATGAAATGGCAATGCATCCGCG 120
DB 61 GGGCTCTATGATGAGCTGGTTACCGCTGTTTACCCATGAAATGGCAATGCATCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTCCGCTACGCTGGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTCCGCTACGCTGGCTGCTGCA 180
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCGAT 240
DB 181 AGCGATGCGCGTAAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCGAT 240
QY 241 GTAGGCCAGGCTGCGATTAATAGTACTTTGAACTGAGTCAATGTTTCAAGAAATAT 300
DB 241 GTAGGCCAGGCTGCGATTAATAGTACTTTGAACTGAGTCAATGTTTCAAGAAATAT 300
QY 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
DB 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAAATGTTGCGTCAAGTT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAAATGTTGCGTCAAGTT 420
QY 421 GGTTTGGCAACACGCCACGCTTAACCAAGTATTTAA 456
DB 421 GGTTTGGCAACACGCCACGCTTAACCAAGTATTTAA 456

```

RESULT 3

AAT74142

ID	Accession	Standard	DNA	Size	BP
XX	AA174142	Standard	DNA	456	BP
XX	AA174142				
AC	25-MAR-2003	(revised)			
DT	29-SEP-1997	(first entry)			
XX					
DE	Salmonella enteritidis 27655-3b	agfa gene			
KW	Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds				
XX					
OS	Salmonella enteritidis				
XX					
PH	Key	Location/Qualifiers			
FT	CDS	1..456			
FT		/*tag= a			
FT		/label= agfa gene fragment			
XX		/transl_except= (pos:367..369,aa:Pro)			
PN	US5635617-A				
XX					
PD	03-JUN-1997				
XX					
PF	26-APR-1994	94US-00233788			
XX					
PR	26-APR-1993	93US-00054452			
XX					
PA	(UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.				
PI	Collinson SK, Kay WW, Doran JL;				
DR	WPI; 1997-309886/28				
XX					
DR	P-PSDB; AAM23570				
XX					
PT	Isolated Salmonella gene agfa - used for diagnosis of Salmonella or enteropathogenic bacteria of the Enterobacteria family				
XX					
XX	Claim 1; Col 19-112; 85pp; English				
XX					
CC	The present sequence represents an isolated agfa gene derived from				
CC	Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide				
CC	diagnostic assays for Salmonella and/or enteropathogenic bacteria of the				
CC	family Enterobacteria. It can also be used to provide proteins and				
CC	antibodies which can be used for assays. The nucleic acid sequence can be				
CC	used to provide probes or primers which can specifically hybridise to				
CC	nucleic acid molecules from greater than 99% of Salmonella strains that				
CC	are pathogenic to warm-blooded animals relative to nucleic acid molecules				
CC	from virtually all other microbial organisms. (Updated on 25-MAR-2003 to				
CC	correct PF field.)				
XX					
SQ	Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;				
XX					
XX	Query Match	88.4%; Score 403.2; DB 2; Length 456;			
XX	Best Local Similarity	92.8%; Pred. No. 2.9e-123;			
XX	Matches 423; Conservative	0; Mismatches 33; Indels 0; Gaps 0;			
QY	1 ATGAAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGGCAGTGCCTGGCT	60			
DB	1 ATGAAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGGCAGTGCCTGGCT	60			
QY	61 GGGCTGTATGATCACTGGTTACCCGTTGTATCCATGAAATGGCAATGATCGGCG	120			
DB	61 GGGCTGTATGATCACTGGTTACCCGTTGTATCCATGAAATGGCAATGATCGGCG	120			
QY	121 CCGAATCAAGTGGAGATTTATAGTACGTTCCGTTAAAGCTGGCGCTTCTTGCA	180			
DB	121 CCGAATCAAGTGGAGATTTATAGTACGTTCCGTTAAAGCTGGCGCTTCTTGCA	180			
QY	181 AGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAAGCGCGCAT	240			
DB	181 AGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAAGCGCGCAT	240			
QY	241 GTAGCGCAGGCTGGGATTAAGTACTGTTGAATGACTGAGAAATGTTTCAGAAATAT	300			

Db	241	GTAGCCAGGGTCCGATPATATGTAATGTAAGTCACTCAGAAATGTTTCAGAAATTAAT	300
Qy	301	GCCACCATCGACCACTGGNACGCTAAAAATCCGATATTACTGTCCGCCAATACGCGGT	360
Db	301	GCACACATCGACCACTGGNACGCTAAAAATCCGATATTACTGTCCGCCAATACGCGGT	360
Qy	361	AATAACGCCCGCGCTGTGTTAATCAGACCCGATCTGATTCAGCGTAAATGGTGGCTACAGTT	420
Db	361	AATAACGCCCGCGCTGTGTTAATCAGACCCGATCTGATTCAGCGTAAATGGTGGCTACAGTT	420
Qy	421	GGTTTGGCAACAACGCCGCTAACGATTTAA	456
Db	421	GGTTTGGCAACAACGCCGCTAACGATTTAA	456
RESULT 4			
AAC64617			
ID	AAC64617 standard; DNA; 456 BP.		
XX	AAC64617;		
AC	26-FEB-2001 (first entry)		
DT	Salmomella enteritidis Agfa DNA sequence SEQ ID NO:1.		
DE	Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;		
XX	vaccine; immune response; immunogen; ds.		
KW	Salmomella enteritidis.		
XX	WO20060102-A2.		
PN	12-OCT-2000.		
PD	05-APR-2000; 2000WO-CA000356.		
XX	05-APR-1999; 99US-0127888P.		
PE	(UUVI-) UNIV VICTORIA.		
XX	White AP, Doran JL, Collison SK, Kay WJ;		
PI	WP1; 2000-672631/65.		
XX	P-PSDB; AAB36341.		
DR	Recombinant agfa gene having a segment replaced by a foreign DNA sequence		
XX	which encodes foreign epitope or antigen, expresses recombinant Agfa		
PT	protein useful for eliciting immune response in animal.		
XX	Disclosure; Page 134; 139pp; English.		
PS	The present invention describes a recombinant agfa gene (I) where a		
XX	segment of the gene has been replaced by a segment of a foreign DNA		
CC	sequence which encodes a foreign epitope or antigen. Also described are:		
CC	(1) use of thin aggregative fimbriae (SFR17/TAF) nucleation depended		
CC	assembly system of strains of Salmomella, Baccherichia coli and		
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant		
CC	Agfa, C8ga and Agfa-homologue fimbria subunits, respectively; (2)		
CC	directing recombination of a recombinant gene into the chromosome of the		
CC	homologous species; (3) directing recombination of a recombinant gene		
CC	back into the chromosome of the homologous species, replacing the native		
CC	copy of that gene; and (4) eliciting an immune response in an animal,		
CC	comprising separating an amino acid polymer comprising a recombinant Agfa		
CC	protein containing a replacement segment or segments of foreign amino		
CC	acid sequence or sequences grown on a Salmomella, E. coli or		
CC	Enterobacteriaceae host cell, from the host cell and introducing the		
CC	polymer into the animal in conjunction with a carrier or diluent. (I) is		
CC	useful for the expression of recombinant Agfa protein which is useful for		
CC	eliciting an immune response in an animal. In a fimbria presentation		
CC	system the heterologous antigens are presented in high numbers (up to		
CC	500,000 copies/cell), the hybrid fimbria protein possesses both the		
CC	immunogenicity and adhesion properties relevant for an efficient live		

White AP, Doran JL, Collison SK
Kearney

361 AATAAGCCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTATGCTGCCTCAGGTT 420

Db 421 GGTTTGGCAACAGCCAGGCTAACAGATATTA 456

RESULT 6
AAC64625 standard; DNA; 456 BP.

XX AAC64625;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#4 DNA sequence SEQ ID NO:17.
XX
XX Salmonella: agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmomella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36349.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/7A) nucleation depended
XX assembly system of strains of Salmomella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmomella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

XX
XX Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 8.5e-107;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGACAGCATTCGTAGTTCTGGCAGTCTTGCT 60
Db 1 ATGAACCTTTAAAGTGGACAGCATTCGTAGTTCTGGCAGTCTTGCT 60
QY 61 GGCCTCTATGATCAGCTGGTTACCCGTTGTTTACCATGAAATGGCAGCATCCGGC 120
Db 61 GGCCTCTATGATCAGCTGGTTACCCGTTGTTTACCATGAAATGGCAGCATCCGGC 120
QY 121 CCGGACTTCACGTTAGACATTTATCAGTACGGTTCCGCTAACGGTCGCTCTGCA 180
Db 121 CCGGACTTCATGATCAGCTGGTTACCCGTTGTTTACCATGAAATGGCAGCATCTGCA 180
QY 181 AGCGATCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCGCAT 240
Db 181 AGCGATCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCGCAT 240
QY 241 GTAGGCCAGGTCGGATTAATAGTACTATTGAACTGACTGAAATGGTTTCAAAATTA 300
Db 241 GTAGGCCAGGTCGGATTAATAGTACTATTGAACTGACTGAAATGGTTTCAAAATTA 300
QY 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATTAATCTGTGGCCAAATACGGCGT 360
Db 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATTAATCTGTGGCCAAATACGGCGT 360
QY 361 AATAACGCCCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGGTCCAGATT 420
Db 361 AATAACGCCCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGGTCCAGATT 420
QY 421 GGTTTGGCAACAGCGCAGGCTAACAGATTTAA 456
Db 421 GGTTTGGCAACAGCGCAGGCTAACAGATTTAA 456

RESULT 7
AAC64628 standard; DNA; 456 BP.

XX AAC64628;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#7 DNA sequence SEQ ID NO:23.
XX
XX Salmonella: agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmomella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36352.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC homologous recombination of a recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention.

XX Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

XX Query Match 77.2%; Score 352; DB 3; Length 456;

XX Best Local Similarity 85.7%; Pred. No. 2.9e-106;

XX Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGGCTCTGGCT 60
 DB 1 ATGAACCTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGGCTCTGGCT 60
 QY 61 GCGCTATGATCAAGCTGGTTACCGGTGTGTTACCCATGAATGGACATGATCCGGC 120
 DB 61 GCGCTATGATCAAGCTGGTTACCGGTGTGTTACCCATGAATGGACATGATCCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTTACCGTTACCG 180
 DB 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTTACCGTTACCG 180
 QY 181 AGCGATGCCCCGTAATTTGAAACGACCTTACCCAGACGGTTATGTTAAACGGCCGAT 240
 DB 181 AGCGATGCCCCGTAATTTGAAACGACCTTACCCAGACGGTTATGTTAAACGGCCGAT 240
 QY 241 GTAGGCCAGGGTGGCGATATAGTACTATGAACTGACCTGAGATGGTTTCAAAATTAAT 300
 DB 241 GTAGGCCAGGGTGGCGATATAGTACTATGAACTGACCTGAGATGGTTTCAAAATTAAT 300
 QY 301 GCCACATCGACCGATGGAACGCTTAAACCTCCGATTTACTGCGCCAAATACGGCGGT 360
 DB 301 GCCACATCGACCGATGGAACGCTTAAACCTCCGATTTACTGCGCCAAATACGGCGGT 360
 QY 361 AATAACCCCGCGTGGTAAATAGACAGCGCATCGATTCAGCGTTATGGTCTCAGGTT 420
 DB 361 AATAACCCCGCGTGGTAAATAGACAGCGCATCGATTCAGCGTTATGGTCTCAGGTT 420
 QY 421 GGTTCGGCAACACGCGCATCGCTTAACCAAGTATTA 456
 DB 421 GGTTCGGCAACACGCGCATCGCTTAACCAAGTATTA 456

RESULT 8
 AAC64622
 ID AAC64622 standard; DNA; 456 BP.

XX AAC64622;
 XX 26-FEB-2001 (first entry)

DE AgfA::PT#1 DNA sequence SEQ ID NO:11.
 XX *Salmonella*: agfA; chromosomal gene replacement; fimbrial epitope;
 KW vaccine; immune response; immunogen; de.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO20060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CM000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UWVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay WW;
 XX WPI: 2000-672631/65.
 XX P-PSDB; AAB56346.
 XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant AgfA
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfA gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
 CC homologous recombination of a recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant AgfA protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention.

XX Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

XX Query Match 77.2%; Score 352; DB 3; Length 456;

XX Best Local Similarity 85.7%; Pred. No. 2.9e-106;

XX Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGGCTCTGGCT 60
 DB 1 ATGAACCTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGGCTCTGGCT 60
 QY 61 GCGCTATGATCAAGCTGGTTACCGGTGTGTTACCCATGAATGGACATGATCCGGC 120
 DB 61 GCGCTATGATCAAGCTGGTTACCGGTGTGTTACCCATGAATGGACATGATCCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTTACCGTTACCG 180
 DB 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTTACCGTTACCG 180

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP, 115 A, 116 C, 118 G, 107 T, 0 U, 0 Other;

Query Match 76.5%; Score 348.8; DB 3; Length 456;
Best Local Similarity 85.3%; Pred. No. 3.3e-105;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGGAGCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGGAGCTCTGCT 60
QY 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATGCTCCGC 120
DB 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATGCTCCGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGTTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGTTAACGCTGCTGCTGCA 180
QY 181 AGCGATGCCGCTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCCGAT 240
DB 181 AGCGATGCCGCTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCCGAT 240
QY 241 GTAGCCGAGGTCGGATTAATAGTACTTATGAACTGACTGCAATGCTTCAAGATAT 300
DB 241 TATATGATGCTGTTACCGGTGTTATCCATGAAATGGCAGCATGCTTCAAGATAT 300
QY 301 GCCACATCGACGATGGAACGCTAAATACTCCGATATTACTGTCGGCAATAGCGCGGT 360
DB 301 GCCACATCGACGATGGAACGCTAAATACTCCGATATTACTGTCGGCAATAGCGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGATT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGATT 420
QY 421 GGTTCGGCAACACGCGCGCTTAACCAATTA 456
DB 421 GGTTCGGCAACACGCGCGCTTAACCAATTA 456

RESULT 12

AAC64627 standard; DNA; 456 BP.

AC AAC64627;
XX 26-FEB-2001 (first entry)
DT XX Agfa::PT36 DNA sequence SEQ ID NO:21.
DE XX Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;
KV vaccine; immune response; immunogen; ds.
XX OS Salmonella enteritidis.
OS Becherichia coli.
OS Synthetic.
XX MO20060102-A2.
PN 12-OCT-2000.
XX 05-APR-2000; 2000MO-CA000356.
PF 05-APR-1999; 99US-0127888P.
PR (UVI-) UNIV VICTORIA.
PA White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.
XX DR

DR P-PSDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

PS The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP1/TF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP, 112 A, 113 C, 125 G, 106 T, 0 U, 0 Other;

Query Match 76.5%; Score 348.8; DB 3; Length 456;
Best Local Similarity 85.3%; Pred. No. 3.3e-105;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGGAGCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGGAGCTCTGCT 60
QY 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATGCTCCGC 120
DB 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATGCTCCGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGTTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGTTAACGCTGCTGCTGCA 180
QY 181 AGCGATGCCGCTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCCGAT 240
DB 181 AGCGATGCCGCTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCCGAT 240
QY 241 GTAGCCGAGGTCGGATTAATAGTACTTATGAACTGACTGCAATGCTTCAAGATAT 300
DB 241 GTAGCCGAGGTCGGATTAATAGTACTTATGAACTGACTGCAATGCTTCAAGATAT 300
QY 301 GCCACATCGACGATGGAACGCTAAATACTCCGATATTACTGTCGGCAATAGCGCGGT 360
DB 301 GCCACATCGACGATGGAACGCTAAATACTCCGATATTACTGTCGGCAATAGCGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGATT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGATT 420
QY 421 GGTTCGGCAACACGCGCGCTTAACCAATTA 456
DB 421 GGTTCGGCAACACGCGCGCTTAACCAATTA 456

RESULT 13

AA064631
ID AAC64631 standard; DNA; 456 BP.

XX AAC64631;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#10 DNA sequence SEQ ID NO:29.

XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.

OS Salmonella enteritidis.

OS Escherichia coli.

XX Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WJ;
XX MPI: 2000-672631/65;
XX P-PSDB; AAB36355.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure: Page 138; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF1/TAFF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC homologous species; (3) directing recombination of a recombinant gene into the chromosome of the
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

XX Query Match: 75.8%; Score 345.6; DB 3; Length 456;

XX Best Local Similarity 84.8%; Pred. No. 3.8e-104;

XX Matches 387; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
XX 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCGGCAGTCTCGGCT 60

Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCGGCAGTCTCGGCT 60
Qy 61 GGCCTCATATGATGAGCTGTTACCCGCTGTTGTTACCATGAAATGGCAGATGATCCGC 120
Db 61 GGCCTCATATGATGAGCTGTTACCCGCTGTTGTTACCATGAAATGGCAGATGATCCGC 120
Qy 121 CCGGACTCAAGCTTGGAGCATTTTACGTAAGGTTGCGTCAAGCGTGGTCTGCGAA 180
Db 121 CCGGACTCAAGCTTGGAGCATTTTACGTAAGGTTGCGTCAAGCGTGGTCTGCGAA 180
Qy 181 AGCGATGCCGTTAAATCTGAAACGACATTTACCGAGCGGTTATGTAACGCGCGAT 240
Db 181 AGCGATGCCGTTAAATCTGAAACGACATTTACCGAGCGGTTATGTAACGCGCGAT 240
Qy 241 GTAGGCGAGGTCGGATTAATGTAATTTGTAAGTCAAGATGTTTCAAAATTAAT 300
Db 241 GTAGGCGAGGTCGGATTAATGTAATTTGTAAGTCAAGATGTTTCAAAATTAAT 300
Qy 301 GCCACCATGACCAAGTGAACGCTTAAACCTCCGATATTACTGTCGGCCATAGCGCGGT 360
Db 301 GCCACCATGACCAAGTGAACGCTTAAACCTCCGATATTACTGTCGGCCATAGCGCGGT 360
Qy 361 AATTAAGCCGCGCTGTTTATCAGACCGCATTTGATTCAGGTTATGTTGCGTCAAGTT 420
Db 361 AATTAAGCCGCGCTGTTTATCAGACCGCATTTGATTCAGGTTATGTTGCGTCAAGTT 420
Qy 421 GGTTTGGCAACACGCGCAGGCTAACCGATTTAA 456
Db 421 GGTTTGGCAACACGCGCAGGCTAACCGATTTAA 456

RESULT 14

AA073066
ID AA073066 standard; DNA; 361 BP.

XX AA073066;

XX 27-AUG-2003 (revised)

XX 25-MAR-2003 (revised)

XX 26-JUN-1995 (first entry)

XX Agfa sequence.

XX Salmonella; Agfa; vaccine; genetic immunization; ds.

XX Salmonella enteritidis.

XX Key Location/Qualifiers

XX CDS 1..359

XX /*tag= a

XX /*note= "Agfa"

XX misc_feature

XX /*tag= d

XX /*note= "TAF5 primer (pair with TAF6)"

XX misc_feature

XX /*tag= b

XX /*note= "TAF3 primer (pair with TAF4)"

XX misc_feature

XX /*tag= e

XX /*note= "TAF6 primer (pair with TAF5)"

XX /*tag= c

XX /*note= "TAF4 primer (pair with TAF3)"

XX WO9425598-A2.

XX 10-NOV-1994.

XX 26-APR-1994; 94WO-IB000207.

XX 26-APR-1993; 93US-00054452.

PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI, 1994-358275/44.
XX P-PSDB; AAR62761.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compans. contg. fibroblast type proteins.
XX
XX Disclosure; Fig 7A; 95pp; English.
XX
XX The DNA encodes the Salmonella enteritidis 27655-3b *typh* mutant strain
CC *agfa* gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to Salmonella in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
XX Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other:
SQ
Query Match 62.3%; Score 284.2; DB 2; Length 361;
Best Local Similarity 99.0%; Pred. No. 8.4e-84;
Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 110 ATGCATCCGCGCCGACCTCAACGTTGACATTATACAGTTCGCTACCGTGGC 169
DB 47 ATAGTTCCGCGCCGACCTCAACGTTGACATTATACAGTTCGCTACCGTGGC 106
QY 170 TTGCTCTGCAAGCGATGCCCGTAAATCTGAAACGACATTACCGAGCGGTTATGTA 229
DB 107 TTGCTCTGCAAGCGATGCCCGTAAATCTGAAACGACATTACCGAGCGGTTATGTA 166
QY 230 ACGGCGCGATGTAGGCGCAGGTCGGAATATAGTACTATTGAAGTGAAGTGAAGTGT 289
DB 167 ACGGCGCGATGTAGGCGCAGGTCGGAATATAGTACTATTGAAGTGAAGTGT 226
QY 290 TCAGAAATTAATGCCCATTCGACCTGAGAACGCTTAAACCTCCGATTTACTGTGGCC 349
DB 227 TCAGAAATTAATGCCCATTCGACCTGAGAACGCTTAAACCTCCGATTTACTGTGGCC 286
QY 350 AATACGCGGTAATTAACGCGCGCTGTATTAATCAGACCGCATCTGATTC 398
DB 287 AATACGCGGTAATTAACGCGCGCTGTATTAATCAGACCGCATCTGATTC 335
RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
AC AAT74141;
XX
DT 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
DB Salmomella enteritidis 27655-3b *typh* mutant *agfa* gene fragment.
XX
KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; da.
XX
OS Salmomella enteritidis.
XX
FH Key Location/Qualifiers
FT CDS 1..360
FT /*tag= a
FT /*label= *agfa*_gene_fragment
FT 16..60
FT /*tag= b
FT /*label= Primer_TAF5
FT 52..69
FT /*tag= c
FT /*label= Primer_TAF3
FT primer_bind complement(103..128)

FT /*tag= d
FT /*label= Primer_TAF6
FT primer_bind complement(294..312)
FT /*tag= e
FT /*label= Primer_TAF4
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX WPI, 1997-309886/28.
XX P-PSDB; AAM23569.
XX
XX Isolated Salmonella gene *agfa* - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX
XX Claim 1; Col 107-110; 85pp; English.
XX
XX The present sequence represents an isolated *agfa* gene fragment derived
CC from Salmonella enteritidis 27655-3b *typh* mutant strain. The nucleic
CC acid can be used to provide diagnostic assays for Salmonella and/or
CC enteropathogenic bacteria of the family Enterobacteria. It can also be
CC used to provide proteins and antibodies which can be used for assays. The
CC nucleic acid sequence can be used to provide probes or primers which can
CC specifically hybridise to nucleic acid molecules from greater than 99% of
CC Salmonella strains that are pathogenic to warm-blooded animals relative
CC to nucleic acid molecules from virtually all other microbial organisms.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other:
Query Match 62.3%; Score 284.2; DB 2; Length 361;
Best Local Similarity 99.0%; Pred. No. 8.4e-84;
Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 110 ATGCATCCGCGCCGACCTCAACGTTGACATTATACAGTTCGCTACCGTGGC 169
DB 47 ATAGTTCCGCGCCGACCTCAACGTTGACATTATACAGTTCGCTACCGTGGC 106
QY 170 TTGCTCTGCAAGCGATGCCCGTAAATCTGAAACGACATTACCGAGCGGTTATGTA 229
DB 107 TTGCTCTGCAAGCGATGCCCGTAAATCTGAAACGACATTACCGAGCGGTTATGTA 166
QY 230 ACGGCGCGATGTAGGCGCAGGTCGGAATATAGTACTATTGAAGTGAAGTGT 289
DB 167 ACGGCGCGATGTAGGCGCAGGTCGGAATATAGTACTATTGAAGTGAAGTGT 226
QY 290 TCAGAAATTAATGCCCATTCGACCTGAGAACGCTTAAACCTCCGATTTACTGTGGCC 349
DB 227 TCAGAAATTAATGCCCATTCGACCTGAGAACGCTTAAACCTCCGATTTACTGTGGCC 286
QY 350 AATACGCGGTAATTAACGCGCGCTGTATTAATCAGACCGCATCTGATTC 398
DB 287 AATACGCGGTAATTAACGCGCGCTGTATTAATCAGACCGCATCTGATTC 335
Search completed: March 15, 2004, 17:51:59
Job time : 246.684 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using BW model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds

(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456
Sequence: 1 atgaactcttaaaagcgc.....ccacgcctaacagatataa 456

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTCUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	403.2	88.4	456	1 US-08-233-788A-58	Sequence 58, Appl
2	284.2	62.3	361	1 US-08-233-788A-56	Sequence 56, Appl
3	33.4	7.3	7766	4 US-09-125-619-3	Sequence 3, Appl
4	31.2	6.8	1491	6 5486473-3	Patent No. 5486473
5	31.2	6.8	2436	4 US-09-540-236-492	Sequence 492, App
6	31.2	6.8	65792	4 US-09-596-002-31	Sequence 31, Appl
7	30.8	6.8	2547	3 US-08-508-761B-1	Sequence 1, Appl
8	30.8	6.8	1664976	4 US-08-916-421B-1	Sequence 1, Appl
9	30.2	6.6	1155	4 US-09-328-352-1775	Sequence 1775, Ap
10	29.6	6.5	972	4 US-09-328-352-2055	Sequence 2055, Ap
11	29.4	6.4	1939	4 US-08-961-527-310	Sequence 310, App
12	29.4	6.4	3494	3 US-09-139-802-200	Sequence 200, App
13	29.4	6.4	3494	4 US-09-659-786-200	Sequence 1006, App
14	29.4	6.4	3494	4 US-09-023-655-1006	Sequence 1006, Ap
15	29.2	6.4	2052	1 US-08-443-104-5	Sequence 5, Appl
16	29.2	6.4	2052	1 US-08-238-130-6	Sequence 6, Appl
17	29.2	6.4	2052	1 US-08-442-859-5	Sequence 5, Appl
18	29.2	6.4	2052	2 US-08-398-489-5	Sequence 5, Appl
19	29.2	6.4	2052	2 US-08-894-772-1	Sequence 1, Appl
20	29.2	6.4	2052	2 US-09-207-844-1	Sequence 1, Appl
21	29.2	6.4	2052	4 US-09-252-509-1	Sequence 5, Appl
22	29.2	6.4	2052	5 PCT-US95-05534-5	Sequence 5, Appl
23	29	6.4	13794	4 US-08-956-171E-54	Sequence 54, Appl
24	29	6.4	1830121	4 US-09-557-884-1	Sequence 1, Appl
25	29	6.4	1830121	4 US-09-643-990A-1	Sequence 1, Appl
26	28.8	6.3	720	4 US-09-328-352-354	Sequence 354, App
27	28.8	6.3	1662	1 US-08-565-386-2	Sequence 2, Appl

28	28.8	6.3	1836	4 US-09-252-991A-15831	Sequence 15831, A
29	28.8	6.3	2055	4 US-09-252-991A-15771	Sequence 15771, A
30	28.8	6.3	2106	4 US-09-252-991A-15891	Sequence 15891, A
31	28.8	6.3	4529	1 US-08-565-386-1	Sequence 1, Appl
32	28.8	6.3	9806	4 US-09-027-169-5	Sequence 5, Appl
33	28.4	6.2	1520	4 US-09-484-970B-30	Sequence 30, Appl
34	28.2	6.2	2116	3 US-09-000-041A-1	Sequence 1, Appl
35	28.2	6.2	2426	3 US-09-028-327-2	Sequence 2, Appl
36	28.2	6.2	2426	4 US-09-571-078A-2	Sequence 2, Appl
37	28.2	6.2	2812	1 US-08-920-812-16	Sequence 16, Appl
38	28.2	6.2	2812	1 US-08-920-827-16	Sequence 16, Appl
39	28.2	6.2	2812	1 US-08-921-177-16	Sequence 16, Appl
40	28.2	6.2	2812	1 US-08-362-577C-16	Sequence 16, Appl
41	28.2	6.2	2812	2 US-08-920-828-16	Sequence 16, Appl
42	28.2	6.2	4973	3 US-09-381-862-6	Sequence 6, Appl
43	28.2	6.2	8078	4 US-09-702-251-3	Sequence 3, Appl
44	28	6.1	906	4 US-09-134-001C-383	Sequence 383, App
45	28	6.1	1665	4 US-09-540-236-1858	Sequence 1858, App

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58

Query Match 88.4%; Score 403.2; DB 1; Length 456;
Best Local Similarity 92.8%; Pred. No. 5e-135;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGACAGATCCGACCAATCGAGTTCTGCGACTGCTTGCT 60
 Db 1 ATGAACTTTTAAAGTGGACAGATCCGACCAATCGAGTTCTGCGACTGCTTGCT 60
 QY 61 GCGCTGTATGATCAGCTGTATACCGCTGTGTATACCATTAATGGACATGCTCGGC 120
 Db 61 GCGCTGTATGATCAGCTGTATACCGCTGTGTATACCATTAATGGACATGCTCGGC 120
 QY 121 CCGGACTGACGTTGAGCTATTATCAGTACGTTCCGCTAAGCTGCGCTGCTGCA 180
 Db 121 CCGGACTGACGTTGAGCTATTATCAGTACGTTCCGCTAAGCTGCGCTGCTGCA 180
 QY 181 AGCATGCTGCTGATGATGAAACGACATTAACCAAGCGGTTATGTTAGTACGCGCGAT 240
 Db 181 AGCATGCTGCTGATGATGAAACGACATTAACCAAGCGGTTATGTTAGTACGCGCGAT 240
 QY 241 GTAGCCAGGCTGCGATTAATGATCTATTGAACTGACTCAGAAATGCTTCAAAATAT 300
 Db 241 GTAGCCAGGCTGCGATTAATGATCTATTGAACTGACTCAGAAATGCTTCAAAATAT 300
 QY 301 GCCACCATGACCAAGTGAACGCTAATAAATCCGATATTCTGCGCCCAATAGCGCGGT 360
 Db 301 GCCACCATGACCAAGTGAACGCTAATAAATCCGATATTCTGCGCCCAATAGCGCGGT 360
 QY 361 AATAAGCGCGCGCTGTTAATGACGCGCATCTGATTCAGCGTAATGCTGCGCTG 420
 Db 361 AATAAGCGCGCGCTGTTAATGACGCGCATCTGATTCAGCGTAATGCTGCGCTG 420
 QY 421 GGTGTTGGCAACACGCGCAGCTAACCAATATTA 456
 Db 421 GGTGTTGGCAACACGCGCAGCTAACCAATATTA 456

RESULT 2 US-08-233-788A-56

/ Sequence 56, Application US/08233788A
 / Patent No. 5635617
 / GENERAL INFORMATION:
 / APPLICANT: Doran, James L.
 / APPLICANT: Kay, William W.
 / APPLICANT: Collinson, Karen S.
 / APPLICANT: Clouthier, Sharon C.
 / TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 / TITLE OF INVENTION: OF SALMONELLA
 / NUMBER OF SEQUENCES: 61
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Seed and Berry
 / STREET: 6300 Columbia Center, 701 Fifth Avenue
 / CITY: Seattle
 / STATE: Washington
 / COUNTRY: U.S.A.
 / ZIP: 98104-7092
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: Patentin Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/233,788A
 / FILING DATE: 26-APR-1994
 / CLASSIFICATION: 435
 / ATTORNEY/AGENT INFORMATION:
 / NAME: King, Joshua
 / REGISTRATION NUMBER: 35,570
 / REFERENCE/DOCKET NUMBER: 920043.403C2
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (206) 622-4900
 / TELEFAX: (206) 682-6031
 / TELEX: 3723836 SEDBANBERY
 / INFORMATION FOR SEQ ID NO: 56:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 361 base pairs

/ TYPE: nucleic acid
 / STRANDEDNESS: double
 / TOPOLOGY: linear
 / MOLECULE TYPE: DNA (genomic)
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: 1..357
 / US-08-233-788A-56

Query Match
 Best Local Similarity 62.3%; Score 284.2; DB 1; Length 361;
 Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 110 ATGATCGCGCGCGGACTCAAGCTTATGACATTTATCAGTACGTTCCGCTAAGCTGCGC 169
 Db 47 ATAGTTCGCGCGCGGACTCAAGCTTATGACATTTATCAGTACGTTCCGCTAAGCTGCGC 106
 QY 170 TTGCTCTGCAAGCGATGCGCGTAAATCGAAAGACCAATTAACCAAGCGGTTATGTA 229
 Db 107 TTGCTCTGCAAGCGATGCGCGTAAATCGAAAGACCAATTAACCAAGCGGTTATGTA 166
 QY 230 ACGGCGCGATGAGCGCAAGGCTGATTAATGATCTATTGAACTGACTCAGAAATGTT 289
 Db 167 ACGGCGCGATGAGCGCAAGGCTGATTAATGATCTATTGAACTGACTCAGAAATGTT 226
 QY 290 TCAGAAATTAATGCGCACATGACAGTGAACGCTAATAAATCCGATATTCTGCGCC 349
 Db 227 TCAGAAATTAATGCGCACATGACAGTGAACGCTAATAAATCCGATATTCTGCGCC 286
 QY 350 AATAAGCGCGGTAATGACGCGCGCTGTTAATGACACCGCATGATTC 398
 Db 287 AATAAGCGCGGTAATGACGCGCGCTGTTAATGACACCGCATGATTC 335

RESULT 3 US-09-125-619-3

/ Sequence 3, Application US/09125619
 / Patent No. 6437116
 / GENERAL INFORMATION:
 / APPLICANT: NORRIS, STEVEN J.
 / APPLICANT: JING-REN, ZHANG
 / APPLICANT: HARDHAM, JOHN M.
 / APPLICANT: HOWELL, JERILYN K.
 / APPLICANT: BARBOUR, ALAN G.
 / APPLICANT: WEINSTOCK, GEORGE M.
 / TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
 / FILE REFERENCE: UTSH:234
 / CURRENT APPLICATION NUMBER: US/09/125,619
 / CURRENT FILING DATE: 1999-01-27
 / NUMBER OF SEQ ID NOS: 48
 / SOFTWARE: Patentin Ver. 2.1
 / SEQ ID NO 3
 / LENGTH: 7766
 / TYPE: DNA
 / ORGANISM: Borrelia burgdorferi
 / FEATURE:
 / NAME/KEY: modified_base
 / LOCATION: (127)
 / OTHER INFORMATION: R = A OR G
 / US-09-125-619-3

Query Match
 Best Local Similarity 7.3%; Score 33.4; DB 4; Length 7766;
 Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 157 GCTAAGCGTGGCTTCTCTGCAAGCGATGCCGTAATTTGAAGACCATTAACCGAG 216
 Db 4432 GTTACGCTGGCGCTGCTGCTGAGCAGATGAGAGAGAGCTTCAAGAGGCTACAAATCCG 4491
 QY 217 AGCGTTATGTAAGCGCGCGCATGTAGCCAGGCTGCGGTAATGATCTATTGAACTG 276
 Db 4492 ATTGCTGCTGCTATTGGAGAGGTAATGAGATGCTGCGGATTTTGTGATGAGATGATG 4551

Qy 277 ACTGAGATGTTTCAGAAATATGSCACCATCGACAGTGGAGGCTAAAACTCCGAT 336
Db 4552 AAGAAGATGATGATGATGCTGCTGCTATGCTTTGAGGGGGGATGCTTAAGATGGAAG 4611
Qy 337 ATTACTGTGGCCCAATACGCGGGTAAATACG 367
Db 4612 TTTCCTGTGAAGAGTAATGATGTGAGAAAG 4642

RESULT 4

5486473-3/c
; Patent No. 5486473
; APPLICANT: FUKITA, HIROYUKI; YOSHIDA, IMAO; TAKAGI, MITSUO;
; MANABE, SADAO; FUKAI, KOSOSUKE
; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
; NUMBER OF SEQUENCES: 16
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,049
; FILING DATE: 09-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 809,255
; FILING DATE: 18-DEC-1991
; APPLICATION NUMBER: 279,685
; FILING DATE: 05-DEC-1988
; APPLICATION NUMBER: 932,419
; FILING DATE: 19-NOV-1986
; SEQ ID NO: 3:
; LENGTH: 1491
5486473-3

Query Match 6.8%; Score 31.2; DB 6; Length 1491;
Best Local Similarity 50.0%; Pred. No. 0.77;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 178 CAAGACGATGCCGTAAATCTGAAACGACCATTCACCGAGCGGTTATGTAAACGGCGCC 237
Db 204 CAATGCTACTGACCGAAGCTAGTAAGTAACTGCGACATCTGCGAGTTGGCTGCTTC 145
Qy 238 GATTTAGGCGGATGCGGATTAATGATCTTGAACGATCTCAAGATGTTTCAGAAT 297
Db 144 CATGTCATCATTTTGAATCAATGTTGGCTTCTTTTGAATTTATGTCACACAACT 85
Qy 298 AATGCCACATCATGACGATGGAACGCTAAAACCTCC 333
Db 84 ATGCCCTTCCAGTACGACATCAACCATGATGCTCC 49

RESULT 5

US-09-540-236-492/c
; Sequence 492, Application US/09540236
; Patent No. 6673810
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709, 2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; NUMBER OF SEQ ID NOS: 2000-04-04
; SEQ ID NO 492
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: M. catarrhalis
US-09-540-236-492

Query Match 6.8%; Score 31.2; DB 4; Length 2436;
Best Local Similarity 53.7%; Pred. No. 1;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

Qy 209 TTACCCAGAGCGGTTATGTTACGCGCCGATGTAGCCAGGGTGGCGATATATGTA 268
Db 1511 TTACCCACATTCATCATATCTTAAAGCAAGTTGGGCGATGTGCTGATTTGAATTACGC 1452

Qy 269 TTGA---ACTGACTCAGATGTTTCAAAATATAGCCACCATGACGATGGAACGCTA 325
Db 1451 TGAATTCATATATCAAAATGCTTCAATTAATTAAGCAGATGAGCTTCTGCAAGAA 1352
Qy 326 AAAATCCGATATTACTGTGCGCCCAATACGCGGGTAAATTAACGCC 369
Db 1391 ACTGTACCTCATGACCATCAACCAAACTTCGGTTGCAATTGCC 1348

RESULT 6

US-09-596-002-31/c
; Sequence 31, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Legace, Robert, E.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 65792
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte template ID No. 6632636 31
; PUBLICATION INFORMATION:
US-09-596-002-31

Query Match 6.8%; Score 31.2; DB 4; Length 65792;
Best Local Similarity 53.7%; Pred. No. 7.1;
Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

Qy 209 TTACCCAGAGCGGTTATGTTACGCGCCGATGTAGCCAGGGTGGCGATATATGTA 268
Db 6819 TTACCCACATTCATCATATCTTAAAGCAAGTTGGGCGATGTGCTGATTTGAATG 6760
Qy 269 TTGA---ACTGACTCAGATGTTTCAAAATATAGCCACCATGACGATGGAACGCTA 325
Db 6759 TGAATTCATATCAAAATGCTTCAATTAATTAAGCAGATGCGCTTGCACAAAG 6700
Qy 326 AAAATCCGATATTACTGTGCGCCCAATACGCGGGTAAATTAACGCC 369
Db 6699 ACTGTACCTCATGACCATCAACCAAACTTCGGTTGCAATTGCC 6656

RESULT 7

US-08-508-761B-1
; Sequence 1, Application US/08508761B
; Patent No. 6027920
; GENERAL INFORMATION:
; APPLICANT: Joliff, Gwenael
; APPLICANT: Guyonvarch, Arnel
; APPLICANT: Purification, Relano
; APPLICANT: Duchillon, Francis
; APPLICANT: Renaud, Michel
; TITLE OF INVENTION: System for Protein Expression and
; TITLE OF INVENTION: Secretion Especially in Corynebacteria
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
; STREET: 400 Seventh St. N.W.
; CITY: Washington D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,761B
FILING DATE: 31-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09652
FILING DATE: 29-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09870
FILING DATE: 02-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,409
REFERENCE/DOCKET NUMBER: P58525NA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 393-5350
TELEFAX: (202) 638-6666
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2547 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Corynebacterium Melassecola
FEATURE:
NAME/KEY: CDS
LOCATION: 420..2390
US-08-508-761B-1

Query Match
Best Local Similarity 6.8%; Score 30.8; DB 3; Length 2547;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Db
280 CAGATGGTTTCAGAAATTAATGCCACCATCGACCACTGAGAGCGTAAACTCCGATATT 339
876 CAGAGTGTGTGACCATTTGACCAATGAGCACTACTACGCGATTAAGACGCCATT 925
340 ACTGTGCGCCATATACGCGGTATTAACGCC 369
936 GTGTGTCTCCCAATCGGTGCGAGAGCTTC 965

RESULT 8
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503728
GENERAL INFORMATION:
APPLICANT: Buit et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-
Patent No. 6503729
FILE REFERENCE: PB375
CURRENT APPLICATION NUMBER: US/08/916,421B
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
LOCATION: (559241)..(559241)

```

```

LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)

```

NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 2055

LENGTH: 972
 TYPE: DNA
 ORGANISM: Acinetobacter baumannii
 US-09-328-352-2055

Query Match
 Best Local Similarity 6.5%; Score 29.6; DB 4; Length 972;
 Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 265 ACTATGACTGCTGAGATGTTTCAAAATATGCCACCATGACGAGGCT 324
 DB 558 ATTAGGACTAATATATATGACCATATATAAGCCAACTGAGCTATGATCT 499
 QY 325 AAAAATCCGATATTA 340
 DB 498 CAATCTCTAGATTA 483

RESULT 11

US-08-961-527-310
 Sequence 310, Application US/08961527
 Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 391
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527
 FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 310:

SEQUENCE CHARACTERISTICS:

LENGTH: 1939 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-310

Query Match

Best Local Similarity 6.4%; Score 29.4; DB 4; Length 1939;

Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 235 GCCGATGTAGCCAGGGTGGGATATAGTACTATTGAAGTGAAGTGGTTTGA 294
 DB 618 GCCAGTGTATGACGGTTAGCGGAAAAAGACTTTACTGAAATTCCTAAGAGTACACCACT 677
 QY 295 AATTAAGCCACATGACGAGTGAAGCTATAAAA 329
 DB 678 AAGGTTCTATCTAGATGAAGTACTGTGAAA 712

RESULT 12

US-09-139-802-200/c

Sequence 200, Application US/09139802

Patent No. 6180084

GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki

APPLICANT: Ruoslahti, Renata

TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing

TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using

FILE REFERENCE: P-1J 3203

CURRENT APPLICATION NUMBER: US/09/139,802

CURRENT FILING DATE: 1998-08-25

EARLIER FILING DATE: 1997-09-10

EARLIER APPLICATION NUMBER: 08/926,914

EARLIER FILING DATE: 1996-09-10

NUMBER OF SEQ ID NOS: 226

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 200

LENGTH: 3494

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (121)..(3024)

US-09-139-802-200

Query Match

Best Local Similarity 6.4%; Score 29.4; DB 3; Length 3494;

Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGTTATGATGTAACGGCCGATGAGCCAGGTTGGATTAATAGTATTG 271
 DB 2531 CCTGGGAGATGAGTTGTCAGTAAACCGGACCGAGTTGGGTTGATTCGGTTATTA 2472
 QY 272 AACTGACTCAGAAATGTTTCAAAATATGCCACATCCGACGAGTGAAGCTAAAACT 331
 DB 2471 TTGGGTTTCCATCCACTCTTGAAGGCGAGACCATCTCCACACTCTGAACT 2412
 QY 332 CCAGATATTACTGTGGCCATACGGCGTAATTAACCGCCGCTGGTAAATCAGCCGAT 391
 DB 2411 CCGTTGAGCAGGCGGTCGATGCAATTAACCTGCTGTACTGCTCCATCAGGTTTCT 2352
 QY 392 CTGATTCACGTAATGTCGTCAGTTGTTTGGCAACAACGCCACGGCTAACCG 450
 DB 2251 GGGATCTCCCTCCAGTTGTGTATTAATTCGAAGTGAATGAAGGCGTGTGACCTG 2293

RESULT 13

US-09-659-786-200/c

Sequence 200, Application US/09659786

Patent No. 6491894

GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki

APPLICANT: Ruoslahti, Renata

TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing

TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using

FILE REFERENCE: P-1J 3203

CURRENT APPLICATION NUMBER: US/09/659,786

CURRENT FILING DATE: 2000-09-11

PRIOR APPLICATION NUMBER: 08/926,914

PRIOR FILING DATE: 1997-09-10

PRIOR APPLICATION NUMBER: 08/710,067

PRIOR FILING DATE: 1996-09-10

NUMBER OF SEQ ID NOS: 226

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 200

LENGTH: 3494

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS
LOCATION: (121) (3024)
US-09-659-786-200

Query Match 6.4%; Score 29.4; DB 4; Length 3494;
Best Local Similarity 45.2%; Pred. No. 5.6;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTAACGGCCCGCATGTAGCCAGGTCGGATATAGTACTATG 271
2531 CCCGCGGAGTAGCGCTTACAGTACGCTGACCGCAGGTTGGGTGATCGGTTATTA 2472
272 AACTGACTCAGATGCTTTCAGAAATATATGCCACCATGACCACTGAACTTAACT 331
2471 TTGGGGTTCTCATCTCCTGCTGTAAGAGCCAGACATCTCTCAGACTTGAACCT 2412
332 CCGATATTACTGTGCGCAATATACCGCGGTATTAACCGCGCTGTTATGACCGCAT 391
2411 CCGTTGAGCAGCGCGCTGCTGATGACATTAACCTGCTGTAAGTCTGATCAGGTTTCT 2352
392 CTGATTCAGCGGTATGTCGCTCAGGTTGGTTTGGCAACAGCCGCTAACCG 450
2351 GGGATCTCCCTCCAGTTGTGTATTTCTGAAATGAAGAGGGGTGTACCTG 2293

RESULT 14

US-09-023-655-1006/C
Sequence 1006, Application US/09023655
Patent No. 6607879

GENERAL INFORMATION:
APPLICANT: Cooke, Benjamin G.
APPLICANT: Susan G. Stuart
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023.655
FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1006:
SEQUENCE CHARACTERISTICS:
LENGTH: 3494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g178535
US-09-023-655-1006

Query Match 6.4%; Score 29.4; DB 4; Length 3494;
Best Local Similarity 45.2%; Pred. No. 5.6;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTAACGGCCCGCATGTAGCCAGGTCGGATATAGTACTATG 271
2531 CCCGCGGAGTAGCGCTTACAGTACGCTGACCGCAGGTTGGGTGATCGGTTATTA 2472
272 AACTGACTCAGATGCTTTCAGAAATATATGCCACCATGACCACTGAACTTAACT 331
2471 TTGGGGTTCTCATCTCCTGCTGTAAGAGCCAGACATCTCTCAGACTTGAACCT 2412
332 CCGATATTACTGTGCGCAATATACCGCGGTATTAACCGCGCTGTTATGACCGCAT 391
2411 CCGTTGAGCAGCGCGCTGCTGATGACATTAACCTGCTGTAAGTCTGATCAGGTTTCT 2352
392 CTGATTCAGCGGTATGTCGCTCAGGTTGGTTTGGCAACAGCCGCTAACCG 450
2351 GGGATCTCCCTCCAGTTGTGTATTTCTGAAATGAAGAGGGGTGTACCTG 2293

RESULT 15

US-08-443-104-5/C
Sequence 5, Application US/08443104
Patent No. 5691162

GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5691162 No. 5691162disk of No. 5691162th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443.104
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-443-104-5

Query Match 6.4%; Score 29.2; DB 1; Length 2052;
Best Local Similarity 49.4%; Pred. No. 4.9;
Matches 76; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

```

Qy 300 TGCACCATGACACGAGTGAACGCTAAATACTCGATATTACTGTGCGCCCAATACGGCG 359
Db 563 TGTAGCCTCAACGGGAGAGAAAGAACATCAACGGTGTCTTGAGAGCGTCGACGGGGT 504
Qy 360 TAATACGCCGCGCTGTATATCAGACCGCATCTGATTCAGGTAATGTGGTCAGGT 419
Db 503 CTTTCTGTCACGCTTGTAAAGAGACCGGGAATCTTGCCCTGTGAAGACTGTTCGGT 444
Qy 420 TGGTTTGGCAACAGCCACGGCTAACCAAT 453
Db 443 AGAGAGAGACCTCGCGCTACGCCCAATCTGTAT 410

```

Search completed: March 16, 2004, 04:37:30
 Job time : 52.6647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using BW model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456
Sequence: 1 atggaactttcaaaagtcgc.....ccacgcctaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.6	7.8	1965	9	US-09-815-242-6264
2	35.6	7.8	1965	12	US-10-282-122A-20332
3	34	7.5	2583	12	US-10-282-122A-19592
4	34	7.5	2886	9	US-09-801-368-131
5	33.6	7.4	549	10	US-09-991-936-1800
6	33.4	7.3	7766	12	US-10-222-566-3
7	33.4	7.3	7766	14	US-10-222-162-3
8	33.4	7.3	7766	14	US-10-143-024-3
9	32.8	7.2	1059	9	US-09-738-626-482
10	32.8	7.2	1113	9	US-09-738-626-481
11	32.8	7.2	3309400	9	US-09-738-626-1
12	32.2	7.1	568	14	US-10-333-631-1
13	32.2	7.1	972	14	US-10-333-631-4
14	32.2	7.1	1083	15	US-10-369-493-33133
15	32.2	7.1	4830	12	US-10-282-122A-27869

16	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
17	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
18	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
19	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
20	31.2	6.8	2394	12	US-10-282-122A-26837	Sequence 26837, A
21	31.2	6.8	9025608	14	US-10-156-761-1	Sequence 1, Appli
22	31	6.8	525	10	US-09-918-995-32144	Sequence 32144, A
23	31	6.8	616	12	US-10-424-599-88945	Sequence 88945, A
24	31	6.8	2751	15	US-10-398-221-35633	Sequence 35633, Ap
25	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
26	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
27	30.8	6.8	1665	15	US-10-369-493-43454	Sequence 43454, A
28	30.8	6.8	2050	14	US-10-086-510-1	Sequence 1, Appli
29	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appli
30	30.6	6.7	2301	9	US-09-815-242-4844	Sequence 4844, Ap
31	30.6	6.7	2301	9	US-09-815-242-5042	Sequence 9042, Ap
32	30.4	6.7	392	12	US-10-282-122A-25231	Sequence 25231, A
33	30.4	6.7	25220	11	US-09-984-429-364	Sequence 364, App
34	30.2	6.6	379	12	US-10-424-599-102524	Sequence 102524, A
35	30.2	6.6	3203	12	US-10-282-122A-24039	Sequence 24039, A
36	30.2	6.6	3222	12	US-10-282-122A-6780	Sequence 6780, Ap
37	30.2	6.6	3222	14	US-10-210-115-19	Sequence 19, Appl
38	30.2	6.6	3222	15	US-10-369-493-24379	Sequence 24379, A
39	30.2	6.6	1601042	15	US-10-027-632-59064	Sequence 59064, A
40	30	6.6	414	14	US-10-156-761-2947	Sequence 2947, Ap
41	30	6.6	1074	12	US-10-282-122A-30822	Sequence 30822, A
42	29.8	6.5	436	12	US-10-424-599-73842	Sequence 73842, A
43	29.8	6.5	815	12	US-10-424-599-72759	Sequence 72759, A
44	29.8	6.5	1394	14	US-10-283-423-13	Sequence 13, Appl
45	29.8	6.5	1394	14	US-10-213-821-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-815-242-6264
Sequence 6264, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trewick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PacSeq for Windows Version 4.0
SEQ ID NO 6264
LENGTH: 1965
TYPE: DNA
ORGANISM: Escherichia coli

.....GAGGAAAGGTCGATGAG 195

304 ACCATCGACCAGTGGAAACGCTAAAACTCCGATATTACTGTGGCCATATCGCGGTAAT 363

Db 196 ATCAACAAGGCGTTCATGTAAGCAACCAAGATATTGAGTGTGACGTGACCGCGCTGT 255
Qy 364 AACGCGCGCTGTTAATCAAGACCGCATCTGA 395
Db 256 ACTGCGCATCTGACTGCTCGCATTTGAAGCTGA 287

RESULT 10

US-09-738-626-481/c
Sequence 481, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 481
LENGTH: 1113
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-481

Query Match 7.2%, Score 32.8; DB 9; Length 1113;
Best Local Similarity 47.2%; Pred. No. 1.9;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 184 GATCCCGTAAATTTGAAGACCAATTAACCAAGCGGTTATGTTACGCGCGCATGTA 243
Db 972 GAGGACGTTTCACCGCAGAGACCAACCACTGCAAGCTCTCCGTTAAGCATCGATGCA 913
Qy 244 GCGCAGGTCGGGTAATAGTACTATTGACTGACTGAGATGTTTCAAGAAATATGCC 303
Db 912 GCGCGTAAAAAGTAAACCATCATCCGCTTCACTTGAGCCTGAGAAAAAGTCGATGAG 853
Qy 304 ACCATGACACAGTGAAGCGTAAAACTCCGATATTACTGTGCGCAATACGCGGTTAT 363
Db 852 ATCAACAAGGCGTTCATGTAAGCAACCAAGATATTGAGTTGAGTGTACCGCGCTGT 793
Qy 364 AACGCGCGCTGTTAATCAAGACCGCATCTGA 395
Db 792 ACTGCGCATCTGACTGCTCGCATTTGAAGCTGA 761

RESULT 11
US-09-738-626-1/c
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 7.2%, Score 32.8; DB 9; Length 3309400;
Best Local Similarity 47.2%; Pred. No. 1.3e+02;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 184 GATCCCGTAAATTTGAAGACCAATTAACCAAGCGGTTATGTTACGCGCGCATGTA 243
Db 451820 GAGGACGTTTCACCGCAGAGACCAACCACTGCAAGCTCTCCGTTAAGCATCGATGCA 451761
Qy 244 GCGCAGGTCGGGTAATAGTACTATTGACTGACTGAGATGTTTCAAGAAATATGCC 303
Db 451760 GCGCGTAAAAAGTAAACCATCATCCGCTTCACTTGAGCCTGAGAAAAAGTCGATGAG 451701
Qy 304 ACCATGACACAGTGAAGCGTAAAACTCCGATATTACTGTGCGCAATACGCGGTTAT 363
Db 451700 ATCAACAAGGCGTTCATGTAAGCAACCAAGATATTGAGTTGAGTGTACCGCGCTGT 451641
Qy 364 AACGCGCGCTGTTAATCAAGACCGCATCTGA 395
Db 451640 ACTGCGCATCTGACTGCTCGCATTTGAAGCTGA 451609

RESULT 12

US-10-333-631-1
Sequence 1, Application US/10333631
Publication No. US20030190651A1
GENERAL INFORMATION:
APPLICANT: Bayer AG
TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
FILE REFERENCE: L10121 Foreign Countries
CURRENT APPLICATION NUMBER: US/10/333,631
PRIOR FILING DATE: 2003-01-23
PRIOR APPLICATION NUMBER: US 60/221,917
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/280,112
PRIOR FILING DATE: 2001-04-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (90)-(90)
OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

Query Match 7.1%, Score 32.2; DB 14; Length 568;
Best Local Similarity 52.6%; Pred. No. 2.1;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 224 ATGTAAAGGCGCCGATGTAGGCCAGGGTGGGATAATAGTACTATTGAAGTGAAGTCTGAGA 283

Db	4229	ATGATAGTAAAGAGACTTCGCGACATTGCTCCGACCAAGATGCTGACCCGATCAGCA	4170
Oy	284	ATGCTTTCAAAATATATCCACCATCGACAGTGAACGCTAAAACTCCGATATTACTG	343
Db	4169	ACGCTGCCACATGCTCGGCTTCGACCAAGGCATCGAGCACTACACCCAGATCGTCG	4110
Oy	344	TCGGCCATAGGGCGGTATATAACGGCGG	372
Db	4109	TCGGTCAAGACGATGTCAAGCGGCTCTCG	4081

Search completed: March 17, 2004, 08:16:08
 Job time : 411.736 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgcgtacacagttatcaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_estbam:*
3: em_estcin:*
4: em_estcmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_estc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_estc3:*
12: gb_estc4:*
13: gb_estc5:*
14: gb_estc6:*
15: em_estfun:*
16: em_estfun:*
17: em_estfun:*
18: em_estfun:*
19: em_estfun:*
20: em_estfun:*
21: em_estfun:*
22: em_estfun:*
23: em_estfun:*
24: em_estfun:*
25: em_estfun:*
26: em_estfun:*
27: em_estfun:*
28: gb_estc1:*
29: gb_estc2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	223	48.9	680 12	B618688 B618688
2	37.4	8.2	484 9	AV430994 AV430994
3	37.2	8.2	772 29	CG862779 CG862779
4	35.8	7.9	484 14	CF089570 CF089570

Result No.	Score	Query Match length	ID	Description
5	35.2	7.7	813 28	B2204853 B2204853
6	34.6	7.6	702 12	BQ441624 BQ441624
7	34.6	7.5	937 12	BQ441625 BQ441625
8	34.2	7.5	296 9	AV058630 AV058630
9	34.2	7.5	523 12	BJ334624 BJ334624
10	34	7.4	811 28	B2469000 B2469000
11	33.8	7.4	480 14	CF372418 CF372418
12	33.8	7.4	481 14	CF414539 CF414539
13	33.8	7.4	509 14	CF605114 CF605114
14	33.6	7.4	594 12	BJ335653 BJ335653
15	33.6	7.4	653 12	BJ336964 BJ336964
16	33.6	7.4	712 14	CF437982 CF437982
17	33.4	7.3	679 29	CC961526 CC961526
18	33.4	7.3	695 28	BH653951 BH653951
19	33.4	7.3	700 28	BH653971 BH653971
20	33.4	7.3	709 28	BH653971 BH653971
21	33.4	7.3	785 28	BH578793 BH578793
22	33.4	7.3	815 28	BH461716 BH461716
23	33.4	7.3	827 28	BH438448 BH438448
24	33.4	7.3	830 28	BH715216 BH715216
25	33.4	7.3	835 28	BH575194 BH575194
26	33.4	7.3	853 28	BH578402 BH578402
27	33.4	7.3	1026 28	B2458784 B2458784
28	33.2	7.3	589 28	B2300068 B2300068
29	33.2	7.3	625 10	BF501499 BF501499
30	33.2	7.3	788 9	AU139321 AU139321
31	33	7.2	456 12	BJ331339 BJ331339
32	33	7.2	524 12	BJ331339 BJ331339
33	33	7.2	547 12	BJ333958 BJ333958
34	33	7.2	553 12	BJ336903 BJ336903
35	33	7.2	569 12	BJ334047 BJ334047
36	33	7.2	574 12	BJ330328 BJ330328
37	33	7.2	607 12	BJ365766 BJ365766
38	33	7.2	625 12	BJ396922 BJ396922
39	33	7.2	629 12	BJ335378 BJ335378
40	33	7.2	635 12	BJ397391 BJ397391
41	33	7.2	645 12	BJ392753 BJ392753
42	33	7.2	670 12	BJ336637 BJ336637
43	32.8	7.2	518 10	BF755518 BF755518
44	32.8	7.2	572 12	BF626794 BF626794
45	32.8	7.2	574 10	BF854273 BF854273

ALIGNMENTS

RESULT 1
LOCUS B618688/c 680 bp mRNA linear EST 01-OCT-2003
DEFINITION B618688 NIBB Mochii normalized Xenopus early gastrula library
ACCESSION B618688
VERSION B618688.1 GI:37256713
KEYWORDS
SOURCE
ORGANISM Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
REFERENCE
1 (bases 1 to 680)
Kikavasa, A., Terasaka, C., Mochii, M., Ueno, N., Shin, I. T. and
Kohara, Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadashi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: shin-i@genes.nig.ac.jp
The information of this clone is available through the following
URL.

FEATURES
source

http://xenopus.nibb.ac.jp.
Location/Qualifiers
1. 680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_id="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 48.9%; Score 223; DB 12; Length 680;
Best Local Similarity 68.0%; Pred. No. 3.1e-58;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGCAGCATTCGACGCAATCGTAGTTCTGGCAGTCTTGGCT 60
DB 593 ATGAACCTTTAAAGTGCAGCATTCGACGCAATCGTAGTTCTGGCAGTCTTGGCT 60
QY 61 GCGCTATGATCAGCTGCTTACCGCTGTTGTTACCCATGAATGCAATGCAATCGGC 120
DB 533 GCGCTATGATCAGCTGCTTACCGCTGTTGTTACCCATGAATGCAATGCAATCGGC 120
QY 121 CCGGACTAAGCTTATGATTTATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 473 CCGGACTAAGCTTATGATTTATGATGATGATGATGATGATGATGATGATGATGAT 180
QY 181 AGCGATGCGCCGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACCGCGCAT 240
DB 413 ACTATGCGCCGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACCGCGCAT 240
QY 241 GTAGCGCGCGGCGGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 353 GTAGCGCGCGGCGGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCCACCATGACAGTGAAGCGTAATAAATCGATGATGATGATGATGATGATGATGAT 294
DB 293 GCTACTCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234
QY 361 AATAACGCGCGCGGTTATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 233 GCGAAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTGTTGGCAACACCGCGCTAATGATGATGATGATGATGATGATGATGATGATGAT 456
DB 173 GCGTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 138

RESULT 2
LOCUS AV430994/c 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TV-1 Porphyra yezoensis cDNA clone
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
1 (bases 1 to 484)
Tabata, S.
Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
DNA Res 7, 223-227 (2000)
20363100
10907854
PUBMED
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers
1. 484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TV-1"
/db_xref="taxon:2788"
/clone="p1027e10_r"
/clone_id="Porphyra yezoensis TV-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 2.4;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 321 GCGTAAACTCCGATATTAATGTCGCGCCATATGCGCGGTAATACCGCGCTGTTAA 380
DB 441 CCGAAGACACAGATGTCGCGCGCTCAACACGCTCAATACGCTGCTGCGCAG 382
QY 381 TCAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 439
DB 381 GCGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 323

RESULT 3
LOCUS CG862779/c 772 bp DNA linear GSS 19-NOV-2003
DEFINITION ZMBB0273B12f ZMBB02 (EcoRI) Zea mays subsp. mays genomic clone
ACCESSION ZMBB0273B12.5, genomic survey sequence.
VERSION CG862779
KEYWORDS GSS.
SOURCE Zea mays subsp. mays (maize)
ORGANISM Zea mays subsp. mays (maize)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
1 (bases 1 to 772)
Zohovetz, V., Fuks, G., Yu, Y., Wing, R. and Messing, J.
Sequencing of the maize genome at PGR (2003)
Unpublished (2003)
Contact: Bharti, A.K.
Dr. Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: 77
Clase: BAC ends
High quality sequence start: 450.
Location/Qualifiers
1. 772
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultiivar="B73"
/sub_species="mays"
/db_xref="taxon:4578"
/clone="ZMBB0273B12"
/lab_host="E. coli DH10B"
/clone_id="ZMBB02 (EcoRI)"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

FEATURES
source

ORIGIN

Query Match 8.2%; Score 37.2; DB 29; Length 772;
Best Local Similarity 51.9%; Pred. No. 3.6;
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 45 TGGCAGTCTCTGCTGGCGCTGTATGATCAGCTGGTACCGCTGTGTTACCATGAAT 104
 DB 202 TGTTCATCAGCTCGCTCCCGCCGCGGGGCTTTCGCTCCCACTTGTCTAACAGAGTGGC 143
 QY 105 GGCACATCATCCGGCCCGGACTCAAGCTTGAAGATTTATGATACGCTTCGGCTAACGC 164
 DB 142 GGCAGTGTCTCCGCAAGACACGCCGATGAGCATCTGTGGTATGTTGGCAGCTCGG 83
 QY 165 TGGCGTCTCTGCAAGAGCATGCCCGTAAATCTGAACGAC 206
 DB 82 TGAGCTGCATTGGAACATTTCAGCAATGTATGTCAGAGAGAC 41

RESULT 4

CF089570 484 bp mRNA linear EST 22-JUL-2003
 LOCUS QHM23504.YG.ab1.QH.M sunflower H. argophyllus Helianthus argophyllus
 DEFINITION CDNA clone QHM23504, mRNA sequence.
 ACCESSION CF089570
 VERSION CF089570.1 GI:3128637
 KEYWORDS EST.

SOURCE

ORGANISM Helianthus argophyllus
 Helianthus argophyllus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; campanulids; Asterales; Asteraceae; Asteroideae;
 Heliantheae; Helianthus.

REFERENCE

1 (bases 1 to 484)
 Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
 Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, U.,
 Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
 Lai, Z., Church, S., Jackson, L. and Bradford, K.
 Letuce and Sunflower ESTs from the Composite Genome Project
 http://composita.ucdavis.edu/
 Unpublished (2002)

JOURNAL

COMMENT Contact: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore Lab
 University of California at Davis (UCD)
 Amundson Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@ucdavis.org [michelmore@ucdavis.edu]
 belongs to contig QH_Ca_Contig2049, see http://cgpdb.ucdavis.edu/
 for details.

Plate: QHM23 row: E column: 04.

FEATURES

Location/Qualifiers

1..484
 /organism="Helianthus argophyllus"
 /mol_type="mRNA"
 /db_xref="taxon:73275"
 /clone="QHM23504"
 /lab_host="E.coli"
 /clone_lib="QH.M sunflower H. argophyllus"
 /note="Vector: pBRCDNA5f1ab; The library was constructed
 from three different sources (seedling, root and leaf) of
 RNA from a single genotype. cDNAs were pooled and
 directionally cloned into a custom medium-copy vector.
 Details of library construction can be obtained at
 http://cgpdb.ucdavis.edu/"

ORIGIN

Query Match 7.9%; Score 35.8; DB 14; Length 484;
 Best Local Similarity 46.0%; Pred. No. 7.4;
 Matches 121; Conservative 0; Mismatches 142; Indels 0; Gaps 0;
 QY 58 GCTGGCGCTATGATCAGCTGTGTTACCGCTGTGTTACCATGAATGCAATGATGCC 117
 DB 323 GCTGGGTTCATTGATTCCTCGATACCGGTCTGCTGAAGATGAATGATACATGCGCT 264
 QY 118 GGGCCGGAATCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 177
 DB 263 GATCTCTGCTGCAATGCGGCTGCTCTCTCAAGGCTTGAATCTGTAGTACGCGCTGAA 204

QY 178 CAAGCGATGCCCGTAATCTGAACGACATTAACCGAGCGGTTATGTATGACGGCC 237
 DB 203 GTGACAAATCATCTATGACAAAGTTTCAATGATCAGAGCTTGTACTCTGTC 144
 QY 238 GATGTAGCCGAGGCTGCGATTAATGATTAATGATGATGATGATGATGATGATGAT 297
 DB 143 CCAGAGGCGGTGTGTCCTAGATTGTGTATGATGATGATGATGATGATGATGATGAT 84
 QY 298 AATGCCACATGACACCACTGGA 320
 DB 83 GAAGCCAAAGATGACCTGTGAA 61

RESULT 5

BZ204853 813 bp DNA linear GSS 11-OCT-2002
 LOCUS CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
 DEFINITION CH230-390D21, genomic survey sequence.
 ACCESSION BZ204853
 VERSION BZ204853.1 GI:23862905
 KEYWORDS GSS.

SOURCE

ORGANISM Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

1 (bases 1 to 813)
 Zhao, S., Shetty, J., Shatsman, S., Tesgaye, G., Geer, K.,
 Shvartsbeyn, A., Gebregorgis, B., Overton, L., Russell, D., Chen, D.,
 Riggs, F., de Jong, P. and Fraser, C.M.
 Rat BAC End Sequences from Library CHORI-230 MboI segment
 Unpublished (1999)
 Other_GSSs: CH230-390D21.TV

JOURNAL

COMMENT Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong [pdejong@mail.cho.org].
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/or ering_information.htm). BAC end
 page: http://www.tigr.org/tcd/bac_ends/rat/bac_end_intro.html
 Plate: 390 row: D column: 21
 Seq primer: T7

FEATURES

Location/Qualifiers

1..813
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SnHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-390D21"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 2"
 /note="Vector: pTARAC1.3; Site 1: MboI; Site 2: MboI;
 CHORI-230 Rat (BN/SnHsd/MCW) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 813;
 Best Local Similarity 49.0%; Pred. No. 16;
 Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
 QY 178 CAAGCGATGCCCGTAATCTGAACGACATTAACCGAGCGGTTATGTATGATGACGGCC 237
 DB 548 CAAGGACCTCCATCAACACAGACACATCAATGATGAAGAAACCTAGGGAAGCA 607
 QY 238 GATGTAGCCGAGGCTGCGATTAATGATTAATGATGATGATGATGATGATGATGATGAT 297

Db 608 TCTGGACACATGGGCACTGGAAAAAATTCCTGACAAACACCAATGGCTATCTCT 667
 QY 298 AATGCCACCATCGACCACTGGAACGCTAAATTCCTGATTTATCTGCGCCAAATACGGC 357
 Db 668 AAGATCAAGATCGACCAATGGATGCTCTTAATCTGCAACCACTTCTGAGCAAGACAC 727
 QY 358 GGTAAATACGCC 369
 Db 728 TGTGTTAGGAC 739

RESULT 6
 BG441624
 LOCUS
 DEFINITION BG441624 702 bp mRNA linear EST 15-MAR-2001
 accession cDNA clone GA_Ea0014A14f, mRNA sequence.
 VERSION BG441624
 KEYWORDS
 SOURCE
 ORGANISM
 Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 1 (bases 1 to 702)
 Henry, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Wing, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 Unpublished (2000)
 Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAATACGACTCACTATGAGG
 High quality sequence stop: 696.
 Location/Qualifiers
 1..702
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0014A14f"
 /tissue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_idb="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source

ORIGIN

Query Match 7 6%; Score 34.6; DB 12; Length 702;
 Best Local Similarity 46.5%; Pred. No. 22;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 181 AGCGATGCCCTTAATCTGAACGACCATTAACCAAGCGGTTATGTAACGCGCCGAT 240
 Db 185 AACATGACGGGAATGCGAATGGGAATGTTGCAACACCAATCAAGACACCAAT 244
 QY 241 GTAGCCAGGGTGGGATTAATGATTAATGTAATGTAATGTAATGTAATGTAAT 300
 Db 245 GACAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 304
 QY 301 GCCACCATGACGAGTGAACGCTAAATTCGATTTATCTGCGCCAAATACGGCGGT 360
 Db 305 GAAATGCTGACAGGACCAACCAATGTAATGTAATGTAATGTAATGTAATGTAAT 364
 QY 361 AATAAGCCGCGCTGTTATCAAGCCGCAATGTAATTCAGGCTATGTTGCTCAGTT 420

Db 365 CAGACGGGAGAAATGCGCACTAGAGTAACACTGACCAATGAATAATGTTGTCAGAT 424
 QY 421 G 421
 Db 425 G 425

RESULT 7
 BG444255
 LOCUS
 DEFINITION BG444255 937 bp mRNA linear EST 15-MAR-2001
 accession cDNA clone GA_Ea0023M14f, mRNA sequence.
 VERSION BG444255
 KEYWORDS
 SOURCE
 ORGANISM
 Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 1 (bases 1 to 937)
 Henry, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Wing, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 Unpublished (2000)
 Contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAATACGACTCACTATGAGG
 High quality sequence stop: 804.
 Location/Qualifiers
 1..937
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0023M14f"
 /tissue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_idb="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source

ORIGIN

Query Match 7 6%; Score 34.6; DB 12; Length 937;
 Best Local Similarity 46.5%; Pred. No. 26;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 181 AGCGATGCCCTTAATCTGAACGACCATTAACCAAGCGGTTATGTAACGCGCCGAT 240
 Db 185 AACATGACGGGAATGCGAATGGGAATGTTGCAACACCAATCAAGACACCAAT 244
 QY 241 GTAGCCAGGGTGGGATTAATGATTAATGTAATGTAATGTAATGTAATGTAAT 300
 Db 245 GACAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 304
 QY 301 GCCACCATGACGAGTGAACGCTAAATTCGATTTATCTGCGCCAAATACGGCGGT 360
 Db 305 GAAATGCTGACAGGACCAACCAATGTAATGTAATGTAATGTAATGTAATGTAAT 364
 QY 361 AATAAGCCGCGCTGTTATCAAGCCGCAATGTAATTCAGGCTATGTTGCTCAGTT 420
 Db 365 CAGACGGGAGAAATGCGCACTAGAGTAACACTGACCAATGAATAATGTTGTCAGAT 424
 QY 421 G 421
 Db 425 G 425

RESULT 8
LOCUS AV058630
DEFINITION AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA clone 1810054102, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 296)
Carrinet, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, D., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Tagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Riken Mouse ESTs
Riken Mouse ESTs
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-rtc.riken.go.jp
Thermolabilization and thermostabilization of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source
location/Qualifiers
1..296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"

ORIGIN

Query Match 7.5%; Score 34.2; DB 9; Length 296;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 251 GTCGGATTAATGCTACTATTAAGCTCAGATGCTTTCGAAATTAATGCCACCATCG 310
Db 13 GTTGAGACCCAGTCTTTCTAGGTCACCAAGCTGGGTAAGAAAGAACATCAACG 72
QY 311 ACCAGTGAAGCGCTAAAGCTCGATATTACTGTGGCCATATACCGCGGTATTAAGCGCG 370
Db 73 TCCAGTGAAGCACAAGAAATGAAGAAATGATTAATCCCAAAAGTTGAATTAAGGTAG 132
QY 371 CGCTGTTAATCAGA 385
Db 133 AGATGCCAAATAGA 147

RESULT 9
LOCUS BJ334624
DEFINITION BJ334624 Dictyostellium discoideum cDNA library, AF Dictyostellium

discoideum cDNA clone dda47006 5', mRNA sequence.
ACCESSION BJ334624
KEYWORDS BJ334624.1 GI:19164754
SOURCE Dictyostellium discoideum
ORGANISM Dictyostellium discoideum
REFERENCE
1 (bases 1 to 523)
Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
Full length cDNA of Dictyostellium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT
Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsnino@genes.nig.ac.jp.

FEATURES
source
location/Qualifiers
1..523
/organism="Dictyostellium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47006"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostellium discoideum cDNA library, AF"

ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 523;
Best Local Similarity 57.1%; Pred. No. 25;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 238 GATGTAGGCCCAAGGCTGCGATATAGTACTATTGAAGTCAAGATGTTTCAAGAAAT 297
Db 398 GATGAAGATGATGATTTGGAGATATGATATATGGAATCAATGGAATGACATTAAT 457
QY 298 AATGCCACCATGACCAAGTGAACGCTAAACCTCCGATATTACT 342
Db 458 AATAACAAACACACACAAATTAATATAAATAAATAATTAAT 502

RESULT 10
LOCUS B2469000/c
DEFINITION B2469000/c 811 bp DNA linear GSS 13-DEC-2002
B2469000/c
Genomic survey sequence.
ACCESSION B2469000
VERSION B2469000.1 GI:26764546
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE
1 (bases 1 to 811)
Town, C.D., Van Aken, S., Uteback, T., Koo, H. and Frazer, C.M.
Whole genome shotgun sequencing of Brassica oleracea
Unpublished (2001)
Other_GSSs: B2469000/c
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.
FEATURES
source
location/Qualifiers
1..811

/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000H3"
/db_xref="taxon:3712"
/clone_lib="BOAO58"
/note="Vector: pHS1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHS1 using BstXI linkers"

ORIGIN

Query Match 7.5%; Score 34; DB 28; Length 811;
Best Local Similarity 48.9%; Pred. No. 37;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 176 TGAAGGATGCGCCGTAATCTGAAGACGACCATTAAGAGCGGTTATGTAACGGC 235
DB 610 TGTAGGCAAGCATTTGTAAACTGTGTCTCCATTTCTACTAGTTTATGACAAAGCT 551
QY 236 CCGATGTAGCGCCGAGGTGGATTAATGACTATTGAAGTCAAGTCAAGTGTTCAGAA 295
DB 550 TTGATCTAATCACTGTGATTTGGACTGCTCTCTGCTTGTCTAGGGAATTTTAAAT 491
QY 296 AATAAGCCAGCATGACCAAGTGAAGCGTAAACTCGATATTACTGTGCGCAATAGC 355
DB 490 ACTTGTCACTCATTAATGAATGAATGAATGACCTGATGATGACATTATCAACCA 431
QY 356 GCGGTA 361
DB 430 AAGATA 425

RESULT 11

CF372418/c

LOCUS CF372418 480 bp mRNA linear EST 27-AUG-2003
DEFINITION CSEC0502808.FL00012 CabSau Normalised Flower Stage 12 (FL00012)
VERSION CF372418
KEYWORDS Vitis vinifera cDNA clone CSEC0502808 3', mRNA sequence.
SOURCE EST
ORGANISM Vitis vinifera
Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; Vitaceae; Vitis.
1 (bases 1 to 480)
Iocco, P., Hua, C., Davies, C. and Thomas, M.R.
Expressed sequence tags from the grapevine cultivar Cabernet
Sauvignon
Unpublished (2003)
Contact: Mark R. Thomas
CSIRO Plant Industry
PO Box 350, Glen Osmond, SA, 5064, Australia
Tel: 61 8 83038600
Fax: 61 8 83038601
Email: Mark.R.Thomas@csiro.au
Seq primer: CCCAGTCACGAGTGTAAACG (M13 Forward)
POLYA=Yes.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENTFEATURES
source

Location/Qualifiers
1..480
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultiVar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSEC0502808"
/sex="Hermaphrodite"
/dev_stage="12 - modified E-L system"
/clone_lib="CabSau Normalised Flower Stage 12 (FL00012)"
/note="Organ: Inflorescence including flowers; Vector:
pZL; Normalised cDNA library from immature inflorescences
at stage 12 of the modified E-L system. Tissue collected
from field grown plants. A description of the modified E-L
system can be found in the paper by B. G. Coombe 'Adoption

of a system for identifying grapevine growth stages'
(1995) Aust. J. Grape and Wine Res. 1: 104-110."

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 480;
Best Local Similarity 49.7%; Pred. No. 31;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 94 ACCCATGAATGGCAGCATGATCCGCCGACCTCAAGCATTTATCAATGACGCT 153
DB 222 ATCATGAGGTGGAACCTGACCTTCCCTCCCAAGAACTGTGCGAGGGCCGCT 163
QY 154 TCCGTTAAGCTGCGCTTCTCTGCAAGGATGCCCGTAATCTGTAAGACCATTAAC 213
DB 162 GCCGCTGCCGCTGCGGCTGCTGCTAGCGAGACAGAGGCTTAAAGCCCATTTCC 103
QY 214 CAGAGCGGTTATGTTAAGCGCGCATGTAGGCCAGGCTGCGATTAATGATAC 266
DB 102 CAGAAAGATCAAGCTTGAGAGAGCATGAGCGGATCAGATGATCTCC 50

RESULT 12

CF414539/c

LOCUS CF414539 481 bp mRNA linear EST 02-SEP-2003
DEFINITION CSEC071B09.POS0036 CabSau Normalised Berry Postveraison Stage 36
(POS0036) Vitis vinifera cDNA clone CSEC071B09 3', mRNA sequence.
VERSION CF414539
KEYWORDS Vitis vinifera
EST.
SOURCE Vitis vinifera
Vitis vinifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; Vitaceae; Vitis.
1 (bases 1 to 481)
Iocco, P., Hua, C., Davies, C. and Thomas, M.R.
Expressed sequence tags from the grapevine cultivar Cabernet
Sauvignon
Unpublished (2003)
Contact: Mark R. Thomas
CSIRO Plant Industry
PO Box 350, Glen Osmond, SA, 5064, Australia
Tel: 61 8 83038600
Fax: 61 8 83038601
Email: Mark.R.Thomas@csiro.au
Seq primer: CCCAGTCACGAGTGTAAACG (M13 Forward)
POLYA=Yes.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENTFEATURES
source

Location/Qualifiers
1..481
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultiVar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSEC071B09"
/sex="Hermaphrodite"
/dev_stage="36 modified E-L system"
/clone_lib="CabSau Normalised Berry Postveraison Stage 36
(POS0036)"
/note="Organ: Fruit with seeds removed; Vector: pZL; A
cDNA library from postveraison fruit at stage 36 of the
modified E-L system. Tissue collected from field grown
plants. A description of the modified E-L system can be
found in the paper by B. G. Coombe 'Adoption of a system
for identifying grapevine growth stages' (1995) Aust. J.
Grape and Wine Res. 1: 104-110."

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 481;
Best Local Similarity 49.7%; Pred. No. 31;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 94 ACCCATGAATGGCAGCATGATCCGCCGACCTCAAGCATTTATCAATGACGCT 153

Db 223 ATCAATGAGGTGGAACCTACCTTCCTCCCTCCCAAGCACTGTGGAGAGGCGCT 164
 Qy 154 TCCGCTACAGCTGCGCTTCTCTGCAAGCGATGCCGTAATTCGAAACACCATACC 213
 Db 163 GCCGTGCGCTGCGCTGCTCTCTGCTAGCAAGACACAGGCGCTAAAGCCACATTTCC 104
 Qy 214 CAGAGCGGTATGTATGTAAGCGCCGATGTAAGCGCGGCGGATTAATAGTAC 266
 Db 103 CAGAAAGATCAAGGCTTGGAAGCAGCAGTACCGGCGATCAGATGATTCTCC 51

RESULT 13
 CF605114 509 bp mRNA linear EST 30-SEP-2003
 LOCUS RADIC01_000700 Grape Root PSPORT1 Library Vitis vinifera cDNA 5',
 DEFINITION mRNA sequence.
 ACCESSION CF605114
 VERSION CF605114.1 GI:37185761
 KEYWORDS EST.
 SOURCE Vitis vinifera
 ORGANISM Vitis vinifera
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; Vitaceae; Vitis.
 1 (bases 1 to 509)
 Moser, C., Segala, C., Fontana, P., Salakhudinov, I., Gatto, P.,
 Pindo, M., Zyprian, E., Toepfer, R., Grandi, M. S. and Velasco, R.
 Expressed sequence tags from different organs of Vitis vinifera
 Unpublished (2003)
 CONTACT: Moser C
 Laboratorio di Genetica Molecolare
 Istituto Agrario di San Michele all'Adige (IASMA)
 via E. Mach 1, San Michele all'Adige (TN), I-38010, Italia
 Tel: 0039-0461-615314
 Fax: 0039-0461-650956
 Email: claudio.moser@iasma.it

FEATURES
 source
 1..509
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultivar="Pinot noir"
 /db_xref="caxon:29760"
 /sex="Hermaphrodite"
 /dev_stage="young root (first year)"
 /lab_host="DH10B"
 /clone_lib="Grape Root PSPORT1 library"
 /note="Organ: root; Vector: PSPORT1; Site_1: Noci; Site_2: Sali"

ORIGIN
 Query Match 7.4%; Score 33.8; DB 14; Length 509;
 Best Local Similarity 49.7%; Pred. No. 32;
 Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 94 ACCCATGAATGGACATGATCGCGCGGAGCTCAAGCTTGAGATTATGATGATGAT 153
 Db 295 ATCAATGAGGTGGAACCTACCTTCCTCCCTCCCAAGCACTGTGGAGAGGCGCT 354
 Qy 154 TCCGCTACAGCTGCGCTTCTCTGCAAGCGATGCCGTAATTCGAAACACCATACC 213
 Db 355 GCCGTGCGCTGCGCTGCTCTCTGCTAGCAAGACACAGGCGCTAAAGCCACATTTCC 414
 Qy 214 CAGAGCGGTATGTATGTAAGCGCCGATGTAAGCGCGGCGGATTAATAGTAC 266
 Db 415 CAGAAAGATCAAGGCTTGGAAGCAGCAGTACCGGCGATCAGATGATTCTCC 467

RESULT 14
 BU335653

LOCUS BU335653 594 bp mRNA linear EST 05-MAR-2002
 DEFINITION BU335653 Dictyostelium discoideum cDNA library, AF Dictyostelium
 discoideum cDNA clone ddas1h06 5', mRNA sequence.
 ACCESSION BU335653
 VERSION BU335653.1 GI:19165783
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 1 (bases 1 to 594)
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 Full length cDNA of Dictyostelium discoideum at the aggregation
 stage
 Unpublished (2002)
 CONTACT: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES
 source
 1..594
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AY4"
 /db_xref="caxon:44689"
 /clone="ddas1h06"
 /sex="mat A"
 /dev_stage="Aggregation stage"
 /clone_lib="Dictyostelium discoideum cDNA library, AF"

ORIGIN
 Query Match 7.4%; Score 33.6; DB 12; Length 594;
 Best Local Similarity 57.1%; Pred. No. 41;
 Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 238 GATGAGGCCAGCGGCGCGATATGACTATGTAACGATCGACATGCTTCAGAAAT 297
 Db 223 GATGAGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 282
 Qy 298 AATGCCACCATCGACCATGGAACGCTAAAACTCCGATATTACT 342
 Db 293 AATAACACACACACACATATATATATATATATATATATATATATATATATATAT 327

RESULT 15
 BU336964 653 bp mRNA linear EST 05-MAR-2002
 LOCUS BU336964 Dictyostelium discoideum cDNA library, AF Dictyostelium
 discoideum cDNA clone ddas5e20 5', mRNA sequence.
 ACCESSION BU336964
 VERSION BU336964.1 GI:19167094
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 1 (bases 1 to 653)
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 Full length cDNA of Dictyostelium discoideum at the aggregation
 stage
 Unpublished (2002)
 CONTACT: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES
 source
 1..653
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"

```

/sfrain="AX4"
/db_xref="taxon:44689"
/clone="ddas5e20"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, Af"

```

ORIGIN

```

Query Match      7.4%; Score 33.6; DB 12; Length 653;
Best Local Similarity 53.9%; Pred. No. 43;
Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 238 GATGTAGGCCAGGTCGGATATAGTACTATTGAACTGACTCAGAAATGTTGAAAT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 GATGAAGATGATGAATTGAGAAATGATATATTGGAATCAATAGAAATGAGACATATAAT 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 AATGCCACCATGACCACTGACGTAACGTAACCTCCGATATTACTGTGGCCATACGGC 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 AATTAACAACACACACATATAATATAAATAATTTATATAAGATGATGAAGAA 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 GGTAATAA 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 AATAAAAA 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 16, 2004, 04:29:07
 Job time : 2232.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignment) 10077.856 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaactcttaaaagcgc.....ccacgcctaccagratraa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

1: GenEmbl:*
2: gb_ba:*
3: gb_hcg:*
4: gb_in:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: gb_wa:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rnd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_ay:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406.4	89.1	456	6	I44909
2	406.4	89.1	2067	1	SEU43280
3	400	87.7	5103	1	STPA2301
4	400	87.7	22411	1	AE008749
5	398.4	87.4	254050	1	AL627269
6	398.4	87.4	301983	1	AE016840
7	392.6	86.1	1048	1	STAF68A
8	282.2	61.9	361	6	I44908
9	280	61.4	2889	1	CSP515700
10	245.2	53.8	2920	1	CFR515701
11	232	50.9	4680	1	BCCSGABDG
12	232	50.9	10346	1	AE000205
13	232	50.9	15047	1	D90741
14	230.4	50.5	456	6	AX814811
15	230.4	50.5	648	1	ECOCGGA
16	220.6	48.4	306358	1	AE016759
17	219	48.0	1711	1	AF275733
18	219	48.0	10190	1	AE005315
19	219	48.0	327773	1	AF002554
20	192	42.1	2883	1	ESAS15702
21	179.4	39.3	230	1	SEU53207
22	163.2	35.8	10370	1	AE015131
23	163.2	35.8	292504	1	AE016981
24	140.2	30.7	437	1	AF237726
25	93.2	20.4	19201	1	D90742
26	64.6	14.2	1212	1	EC0131756
27	48.2	10.6	78	6	AX814809
28	39.4	8.6	10999	1	AE002536
29	39.4	8.6	349980	6	AX044034
30	37.8	8.3	349061	1	NMA222491
31	37	8.1	9955	1	AE002504
32	37	8.1	349980	6	AX044033
33	36.8	8.1	126062	2	AP006135
34	36.6	8.0	110000	2	AC142602_0
35	36.6	8.0	305730	2	AC114347
36	36.6	8.0	311823	1	AC095362
37	36.4	8.0	10798	1	AE004604
38	36.4	8.0	110000	2	AC110929_4
39	36.4	8.0	181996	2	AC108614
40	36.4	8.0	212174	2	AC105896
41	36.4	8.0	305724	2	AC128763
42	36.2	7.9	1797	3	LEIMSP52A
43	36.2	7.9	1860	3	LIGP63GEN
44	36.2	7.9	2104	3	LD0495002
45	36.2	7.9	2106	3	LD0495006

ALIGNMENTS

```

RESULT 1
LOCUS      I44909
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION  I44909
VERSION     I44909.1
KEYWORDS   GI:2469622
SOURCE      Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 456)
AUTHORS   Doran,J.L., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.
TITLE      Methods and compositions comprising the agfa gene for detection of
           Salmonella
JOURNAL    Patent: US 5635617-A 58 03-JUN-1997;

```



```

gene
3344..3376
/gene="csbBA"
-35_signal
3344..3349
/gene="csbBA"
-10_signal
3371..3376
/gene="csbBA"
gene
3469..3924
/gene="csbB"
CDS
3469..3924
/gene="csbB"
/function="nucleator"
/codon_start=1
/evidence=experimental
/product="CsbB protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translation="MKNLPLPMLTILIGAPILATNTYDLARSEYFPAVNEISKSPFN
QAALIGQVGTDRSARVRQSKLSVISOEGNNRRAVDAGNYNPAVLEQTNANDA
SISQSAVGNASALIKSGNNKANTTGTGTAVVVKOSHMAIRVQR"
3469..3531
/gene="csbB"
3532..3921
/gene="csbB"
/product="CsbB protein"
/function="nucleator"
3966..4421
/gene="csbA"
3966..4421
/gene="csbA"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/translation="MKLVAAFAAIVVSGSALAGVVPOMGGGNGHNGCGSSPDST
LSTVYQGSAAALALQSDARKSETTITGSGNGADVGCGADNSTIELTQNGFNNAT
IDQWNAKNSDITVGYGNNALVNGTASDSVWVRQVGFNNATANQY"
3966..4025
/gene="csbA"
4026..4418
/gene="csbA"
/product="CsbA protein"
/function="fiber subunit"
4483..4809
/gene="csbC"
4483..4809
/gene="csbC"
/codon_start=1
/translation="MHTLILALSNQITTTTQGDITVPIQVTLNBPCCVQVQIL
SVRDVGQSHQKQKLTSLPANQPIELSLVNISSEDSVKIITVSDGQSLHSQ
WPSAQ"
4483..4533
/gene="csbC"
4534..4806
/gene="csbC"
/product="CsbC"
4940..5103
/gene="ORF103"
4940..5103
/gene="ORF103"
/codon_start=1

```

ORIGIN

Query Match 87.7%; Score 400; DB 1; Length 5103;
 Best Local Similarity 92.3%; Pred. No. 1,4e-101;
 Matches 421; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGAGCATTCGACGATCGTAGTTCTGCGAGTGTGGCT 60
   3966 ATGAACCTTTAAAGTGGAGCATTCGACGATCGTAGTTCTGCGAGTGTGGCT 60
DB 61 GCGCTCGTCCACAAATGGGGCGGCGGCTAATCATTAACGGCGGCAATAGTCCGAC 120
   4026 GCGCTCGTCCACAAATGGGGCGGCGGCTAATCATTAACGGCGGCAATAGTCCGAC 120
QY 121 CCGACTATGATCAGCTGTTACCCGCTGTTGTAACCATGAATAGGCATGCACTGCA 180
   4086 CCGACTATGATCAGCTGTTACCCGCTGTTGTAACCATGAATAGGCATGCACTGCA 180
DB 181 AGCGATGCCGCTAATCTGAAACGACATTACCCAGCGGTTATGTATAGCGCGCAT 240
   4146 AGCGATGCCGCTAATCTGAAACGACATTACCCAGCGGTTATGTATAGCGCGCAT 240
QY 241 GTAGGCCAGGTCGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
   4206 GTAGGCCAGGTCGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 301 GCCACCATGACGACATGGAAGCGTAAATCTCCGATTAATGTAATGTAATGTAAT 4265
   4266 GCCACCATGACGACATGGAAGCGTAAATCTCCGATTAATGTAATGTAATGTAAT 4265
QY 361 AATTAACGCCGCGCTGTTAATCAAGCCGATCTGATTCAGCGGTAATGTAATGTAAT 420
   4326 AATTAACGCCGCGCTGTTAATCAAGCCGATCTGATTCAGCGGTAATGTAATGTAAT 420
DB 4326 AATTAACGCCGCGCTGTTAATCAAGCCGATCTGATTCAGCGGTAATGTAATGTAAT 420
QY 421 GGTTTGGCAACACGCGCAACGCGTAAACGATTAAT 456
   4366 GGTTTGGCAACACGCGCAACGCGTAAACGATTAAT 456
DB 4366 GGTTTGGCAACACGCGCAACGCGTAAACGATTAAT 456

```

RESULT 4

AE008749

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TTLBL

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

CONSTRM

TITLE

JOURNAL

22411 bp DNA linear BCT 23-APR-2003
 Salmonella typhimurium LT2, section 53 of 220 of the complete genome.
 AE008749 AE006468
 AE008749.1 GI:16419641
 ORGANISM
 Salmonella typhimurium LT2
 Salmonella typhimurium LT2
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.
 1 (bases 1 to 22411)
 McCllland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Latreille, P., Courtney, L., Portnoy, L., S. Ali, J., Danne, M., Du, F.,
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Grewel, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*
 LT2
 Nature 413 (6858), 852-856 (2001)
 21534948
 11677609
 2 (bases 1 to 22411)
 The *Salmonella typhimurium* Genome Sequencing Project
 Direct Submission
 Submitted (23-MAR-2001) Genome Sequencing Center, Department of

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI33283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER, <http://www.tigr.org/softlab/glimmer/glimmer.html> and Genemark; <http://opal.biology.gatech.edu/Genemark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcolCyc; <http://ecocyc.org/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and RegionDB;

http://kinch.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

FEATURES

SOURCE

location/Qualifiers

1..22411

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

434..1308

/gene="phoH"

/note="synonym: STM1126"

434..439

/gene="phoH"

/note="putative RBS for phoH; RegulonDB:STMSIH001398"

454..1308

/gene="phoH"

/note="similar to E. coli phoB-dependent, ATP-binding pho regulation component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"

aa), 92% identity in aa 71 - 354"

/codon_start=1

/transl_table=1

/product="PhoB-dependent ATP-binding pho regulon component"

/protein_id="AAL20057.1"

/db_xref="GI:16419642"

/translation="MCANRVQENAGDVLQKLVSQKQAVIKAREAKRLRDS RSHKQREESVTSIAKAAABALIHQDVERIIVTRPVLOADDLPGDIAEKAPRPV YDVLKRLGASPMQYCLREIGKVIETAPFAVAVYILDEAQNVAQOKMF LTRIGENVTVINQDITQCDLPRGVRSLDALRFRFEDENVGIVHFKDCCVRSALC QRTLHYS"

/complement(1414..2303)

/gene="STM1127"

/complement(1414..2295)

/gene="STM1127"

/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%

identity in aa 15 - 264"

/codon_start=1

/transl_table=1

/product="putative transcriptional regulator"

/protein_id="AAL20058.1"

/db_xref="GI:16419643"

/translation="WEPQPRKPKKILDTGAMOKSITRASQRIAYITLAPROVTO SSIAQRETOAGEATVIRFCRTLGKGFOPKMDLAIELATTESDDSPLLDEVSE SDDAHAIQKLONTISNVLSFTLNLLDMQVIGVVDALRHCHSVYIPGVSGGITLAD MKHKLMRIGLRGDAVSNHFMVQMTLLKAGDVAMGVSHSGTSPETVHSLRLAQGA

RBS

TTVAITHNIGSPICEADPCLINGNRQGLQODSIGTKAQLFVFDLLYTLLVOSSPE QARESKLRPMNALDMTK"

/complement(2298..2303)

/gene="STM1127"

/note="putative RBS for STM1127; RegulonDB:STMSIH001399"

/complement(2580..4088)

/gene="STM1128"

/complement(2580..4076)

/gene="STM1128"

/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%

identity in aa 7 - 478"

/codon_start=1

/transl_table=1

/product="putative sodium/glucose cotransporter"

/protein_id="AAL20059.1"

/db_xref="GI:16419644"

/translation="MITHSPGIVNVLVFGYLLAMLVGVFSRRQKTADYDFRGGR VPGMAQVSPATLSSITFMSIPAKFTSDMTFIIQGYLAIALIPLVFVPIPERK LKMSAYVLEAPDVRCLRFLPASMSPMLFHIGRAIITPLVLAIRPAILDPVLYL

LISVNCITTMGIGEVITDVIQGLLSSAILIFIVTLKQGGIDEIFVTQQA DKFPATQFHSWSTESTVPLMIGLGFANIQFTASQVQVRYVTDSEBTKTLT

NAKLVAVIPVEFPAGSALFVYQDHPQLAGFNTGILPLFVTEMPVGIAGLIA AIFPAQSSISSINSISSCNSDIYORLSHKRTPEPRMKIATKVLIVAGLISAS

VMLVMADESEIMDAFNSLIGLGMGPMGLFMLGIFPKRANSAVIGIISVITVIGA RYADINPFYVGSLSVTSVYIFAPLFAFPAPPLTDERPERVTL"

/complement(4083..4088)

/gene="STM1128"

/note="putative RBS for STM1128; RegulonDB:STMSIH001400"

/complement(4413..5099)

/gene="STM1129"

/complement(4413..5093)

/gene="STM1129"

/note="similar to E. coli putative enzyme (AAC76255.1);

Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -

226"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20060.1"

/db_xref="GI:16419645"

/translation="MSLLARLEQSVHENGVLIVGCPVPSPMDKPEIIVAAQAAS AGFASVREGENETRVPHLSVPIIGIKDLSGVPRTIPYLOVDVALAOGADII

AFDASFSRPPVDISLITRTRHGLMAADCSYNEGSKGKIEFTDITSCVGTG ITVPDPLAVTQLSHACRYIARGRTVTPALANALBHGMAVTVGSATIRIEHIQ

WFSHAVKX"

/complement(5094..5099)

/gene="STM1129"

/note="putative RBS for STM1129; RegulonDB:STMSIH001401"

5599..6759

/gene="STM1130"

5599..6759

/gene="STM1130"

/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%

identity in aa 40 - 403"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20061.1"

/db_xref="GI:16419646"

/translation="MGMQKPKKQMTMLCSPVAITTSQVATLDPIDPEPLKNGS AIDNGVTVVGLGATGTSWKRIDLKKQKQWERRKSPGQARBSQSVFPLDELVYRG

GVGKNSQESPLQVSDVYKTSVKNATQKQVDTISPVGLTGHTGVKLSTWVLLITGVN EHIIPDKYFIDIAAADADESKNKYIVYFNKPADYFNNKLVFIYNKAKENTKRAQBL

PGAGTAGSSVYGNPFLMLINGELKPGIPLRTDVIIRAMVDNKLFWLKNSSQLPSPBQ

QOEGIAQFSGSHVGLVGGANFPAGKQVTKYKSHGKIKKRWDEVYGLINGH

QOEWKMKQKPGISGVSYGDBVFLIGENAKKQVSVYSFTMRDGNLTK"

6793..7497

/gene="STM1131"

6793..6798

/gene="STM1131"

/note="putative RBS for STM1131; RegulonDB:STMSIH001402"

CDS

6805..7497
 /gene="STM1131"
 /note="similar to E. coli orf, hypothetical protein (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 268 identity in aa 4 - 239"
 /codon_start=1
 /transl_table=1
 /product="putative outer membrane protein"
 /protein_id="AAL20062.1"
 /db_xref="GI:16419647"
 /translation="MKINKYLGMVSPFASSTYQAAATLDYRHYADRTINQRIAI
 IEKPNIGIFVDASVSGVDGDDKSLDNLVNAIEICSYNYKTDFVLOPEFI
 RESGPDITIKYPLKRGYNPDGSGVMGRYRYPDKATNATISDEKINRDTYIGVF
 DELKLEYNFTWMDSDQIKFDKNTNIEHNVALMKLNKSPYVEGVNNAVRRNTDER
 ORYRVGLQVH"
 7768..9060
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STWS1H001403"
 7768..7773
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STWS1H001403"
 7780..9060
 /gene="STM1132"
 /note="similar to E. coli putative transport protein (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 576 identity in aa 25 - 418"
 /codon_start=1
 /transl_table=1
 /product="putative sugar transport protein"

Query Match 87.7%; Score 400; DB 1; Length 22411;
 Best Local Similarity 92.3%; Pred: No. 1,6e-101;
 Matches 421; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGCGACATTCGCGACATGCTGATGCTTCTGCGAGTCTCTGCT 60
 DB 17769 ATGAACCTTTAAAGTGCGACATTCGCGACATGCTGATGCTTCTGCGAGTCTCTGCT 17828
 QY 61 GGGGTGCTTCCAAATGGGCGCGCGGCTAATCATATACGCGCGCGCAATAGTCCGCGC 120
 DB 17829 GGGGTGCTTCCAAATGGGCGCGCGGCTAATCATATACGCGCGCGCAATAGTCCGCGC 17888
 QY 121 CCGGACTATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAAATGCCATCATCTGCA 180
 DB 17889 CCGGACTATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAAATGCCATCATCTGCA 17948
 QY 181 AGGATGCGCGGTAATCTGAAACGACATTACCGACGCGGTTATGTTACGCGCGCGAT 240
 DB 17949 AGGATGCGCGGTAATCTGAAACGACATTACCGACGCGGTTATGTTACGCGCGCGAT 18008
 QY 241 GTAGGCGAGGGTGGGATTAATAGTACTTGAATGACTGATGCTTTCAGAAATTAAT 300
 DB 18009 GTAGGCGAGGGTGGGATTAATAGTACTTGAATGACTGATGCTTTCAGAAATTAAT 18068
 QY 301 GCCACATTCGACGATGGAACGCTTAAAACTCCGATATTACTGCGCCAAATACGGCGGT 360
 DB 18069 GCCACATTCGACGATGGAACGCTTAAAACTCCGATATTACTGCGCCAAATACGGCGGT 18128
 QY 361 AATAACGCGCGGCTGTTAATCAGACGCGCATCTGATTCGCGCATATGCTGCGTCAAGTT 420
 DB 18129 AATAACGCGCGGCTGTTAATCAGACGCGCATCTGATTCGCGCATATGCTGCGTCAAGTT 18188
 QY 421 GCTTTTGCAACAACGCCACGCGGTACACAGATTTAA 456
 DB 18189 GCTTTTGCAACAACGCCACGCGGTACACAGATTTAA 18224

RESULT 5
 LOCUS AL627269 254050 bp DNA linear BCT 04-JUL-2003
 DEFINITION *Salmonella enterica* serovar Typhi (*Salmonella typhi*) strain CT18,
 complete chromosome; segment 5/20.
 ACCESSION AL627269 AL513382
 VERSION AL627269.1 GI:16502231

KEYWORDS
 SOURCE
 ORGANISM
 REFERENCES
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 JOURNAL
 COMMENT
 FEATURES
 source

Salmonella enterica subsp. enterica serovar Typhi
 Salmonella enterica subsp. enterica serovar Typhi
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.
 1 (bases 1 to 254050)
 Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,
 Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G.,
 Sebatian, M., Baker, S., Basham, D., Brooks, R., Chillingworth, T.,
 Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N.,
 Farrar, J., Feltham, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S.,
 Jagels, K., Kirov, A., Larsen, T.S., Leather, S., Moule, S., O'Garra, P.,
 Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
 Stevens, K., Whitehead, S. and Barrall, B.G.
 Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18
 Nature 413 (6858), 848-852 (2001)
 2 (bases 1 to 254050)
 Parkhill, J.
 Direct Submission
 Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
 Sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
 Hinxton, Cambridge CB10 1SA, UK
 E-mail: parkhill@sanger.ac.uk
 Notes:
 Details of S. typhi sequencing at the Sanger Centre are available
 on the World Wide Web.
 (URL: http://www.sanger.ac.uk/Projects/S_typhi/).
 Location/Qualifiers
 1..254050
 /organism="Salmonella enterica subsp. enterica serovar
 Typhi"
 /mol_type="genomic DNA"
 /strain="CT18"
 /db_xref="taxon:90370"
 181..2793
 /gene="STY1078"
 /note="synonym: pepN"
 181..2793
 /gene="STY1078"
 /note="orthologue of E. coli pepN (AMPN ECOLI); Pasta hit
 to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
 /codon_start=1
 /transl_table=1
 /product="aminopeptidase N"
 /protein_id="CAD08183.1"
 /db_xref="GI:16502232"
 /db_xref="GOA:Q82770"
 /db_xref="SPTREMBL:Q82770"
 /translation="MTQOPQAKYRHHYRVPDYQITDIEFLDPAEKVTVAISOAVR
 HGADPADLRDGDLLVSIHVNDAPWYKKEEGALLISDLPBPTLRIVNEISPA
 NTALEGLYQSGDALCTQCEAEGRHITWLDSPDVARPTTIDAKSKYPLISNGN
 RVAGLELERNHWOMODPEPKPCYFALVADDPVLRPTTTRGREAELVYDRG
 NLRAPAMNLSLNKSMKMDPEPRGLEVDIDITVAVDFFNMGAMENGLNIPNSKY
 LARDTRADCDYDIERVIGHYTHWNTGRVTCDFQLSKBELTVFRDSESSDL
 GSPAVNRISVTRRGLOFAEDASPMVAHIREDKYIENNNFYTLTYVEGAEVIRMIH
 TILGSENFQKGMQLYFERHDGSAATCDPVOQAMEASVNDLSHFRWVSOGSTPIVTV
 KDYNPEBQETFDYNYFQVPALCEFSAPVKELYKMSDQOCTPMHMAHNDREBRND
 VLVNTGABQETFDYNYFQVPALCEFSAPVKELYKMSDQOCTPMHMAHNDREBRND
 IAELFEVIDPIAIAOVREALTPRTLAELADEFLATYNNHLDKRVHGDIGKTLRN
 OGELVMDKRFIIQSTSPANVLETVRSLLKGRSFMSNPNRIRSLIGAFAGSNPAFH
 AODGSGYQPLVEMTLDNSRNQVARSRLIEPLIRKRVDDRKQEMRAALBQLGLEN
 LSGDLERYTKLA"
 532..1332
 /gene="STY1078"
 /note="Pfam match to entry PF01433 peptidase_M1, Peptidase
 family M1, score 245.10, E-value 1.9e-84"
 1060..1089
 /gene="STY1078"
 misc_feature
 misc_feature
 misc_feature


```

QY      301  GCCACCATGACAGTGAACGCTAAACCTCCGATTACTGTGCGCCAAATACGGCGGT
Db      89206 GCACACATGACAGTGAACGCTAAACCTCCGATTACTGTGCGCAATACGGCGGT 89265
QY      361  AATAACGCGCGCGTGTATATCAGACCGCATCTGATTCAGCGTAAATGTCGACAGTT
Db      89266 AATAACGCGCGCGTGTATATCAGACCGCATCTGATTCAGCGTAAATGTCGACAGTT 89325
QY      421  GGTTTGGACAGACGCGCGCTAACCACTTTAA 456
Db      69326 GGTTTGGACAGACGCGCGCTAACCACTTTAA 89361

RESULT 6
LOCUS   AE016840/c 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmoneella enterica subsp. enterica serovar Typhi Ty2, section 7 of
ACCESSION AE016840 AE014613
VERSION   AE016840.1 GI:29137797
KEYWORDS
SOURCE
ORGANISM Salmoneella enterica subsp. enterica serovar Typhi Ty2
REFERENCE Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G. F., Rose, D. J.,
AUTHORS Burland, V., Kodyalmani, V., Schwartz, D. C. and Blattner, F. R.,
TITLE Comparative Genomics of Salmoneella enterica Serovar Typhi Strains
JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)
MEDLINE 22531367
PUBMED 12644504
REFERENCE 2 (bases 1 to 301983)
AUTHORS Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G. F., Rose, D. J.,
TITLE Direct Submission
JOURNAL Submitted (25-SEP-2002) Laboratory of Genetics, University of
FEATURES
source Location/Qualifiers
1. .301983
/organism="Salmoneella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
/complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
/complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
/notes="corresponds to STY1225 from Accession AL513382:
Salmoneella typhi CT18"
CDS
gene
/locus_tag="t1735"
/notes="corresponds to STY1225 from Accession AL513382:
Salmoneella typhi CT18"
/transl_table=1
/codon_start=1
/product="transposase for insertion sequence element
IS200"
/protein_id="AA069359.1"
/db_xref="GI:29137798"
/translation="MGDEKSLAHPRNMCKHYIVAPRRROAFYGEKRAVSTIRKL
CEWNRVILAEACCAHIMLEIPPMGVSVSMGYLKSSIMLYEGCGDLKFTYRN
REFWRCGYVDYIKNTAKIODYIKQLBEDKMGEDLSIPYSPPTGRK"
/gene="fliG"
/locus_tag="t1736"
/complement(844..1797)
/gene="fliG"
/locus_tag="t1736"
/complement(844..1797)
/notes="corresponds to STY1223 from Accession AL513382:
Salmoneella typhi CT18"
CDS
gene
/locus_tag="t1737"
/complement(1812..3473)
/gene="fliG"
/locus_tag="t1737"
/complement(1812..3473)
/notes="corresponds to STY1222 from Accession AL513382:
Salmoneella typhi CT18"
/transl_table=1
/codon_start=1
/product="flagellar hook-associated protein 1"
/protein_id="AA069361.1"
/db_xref="GI:29137800"
/translation="MSLLINHAMSGLNAQAALNTVSNINNNVAVAGYRQTTILAO
NSTLGGWINGVYSGVREYDAFTITQLRGAQNSGGLTRREQMSKIDNLADK
SSLSGSLQSFSTLQTLVSNADPAAQALIGKAEGLVNOFKTIDYLRDQKQVNI
AIGSSVAQINNYAKOIANINDOISRTYVGAGASPNLIDRDLVSLNKGIVENS
TSGSLGLTFRSODLDQTRNTLIGQLAFAAFAENQHTGVADAGNKGDFEISPV
VSNSSNNAKTVSLTAKVVDSTKQVATLDFDGTQVTRTADNTFTFKADG
LEIDGLKTVGAQKNDSEFLKPVSNALVMMNVATVTEAEIAMAASBKLDPDVTD
SNRNGQALDLDQNSNVVAGNKTFNDAAVATLVSDVGNKSTLSSSTTQANVAVQLK
COQSVGVLDDEYGNLRYOQYVLANQVQITNALFDALNIR"
/gene="fliG"
/locus_tag="t1738"
/complement(3538..4488)
/gene="fliG"
/locus_tag="t1738"
/complement(3538..4488)
/notes="corresponds to STY1221 from Accession AL513382:
Salmoneella typhi CT18"
/transl_table=1
/codon_start=1
/product="flagellar protein fliG"
/protein_id="AA069362.1"
/db_xref="GI:29137801"
/translation="MIGDGKLLASAAWDAQSLNELKAKAGDPANIRPARVEGMF
VQMLKSMREALPKDGLFSSDQTRLYTSWVDOOIACQMTKGLGLADMMVKOMTGGQ
TMPADDAPOVPLKSLFETVNSYONQALTOIVRAIFRTDPSDGLSGSKDFLALS
LPARLASQSGVPHHLLAOLALSGGQKQILRENGEBSYVAVGKATASWKGFLPTE
ITTYEENGAKKVAKAFRVYSYLELSDYVALLRNRYAVATTAATAEACQAVALO
NAGYATDPNYPARKIASMIQOLKAMSEKVSQTSANLNNLF"
/gene="fliG"
/locus_tag="t1739"
/complement(4488..5591)
/gene="fliG"
/locus_tag="t1739"
/complement(4488..5591)
/notes="corresponds to STY1220 from Accession AL513382:
Salmoneella typhi CT18"
/transl_table=1
/codon_start=1
/product="flagellar P-ring protein precursor"
/protein_id="AA069363.1"
/db_xref="GI:29137802"
/translation="MYVFKALAGIVLAVATLAHAERIRDLTSVGVRENSLIGVGLV
VGLDGTQDTQTPFTTQTLNMLSLQGLIVPTGNMOKVAVAAVTVASPPAROG
QITDVVSSMGNAKSLRGGLTMLPLKGVDSVVVALAGNLLVGGASAGGSSQVQV
QNLNGVITNGALIERELPTQFGAGNTINLQNDENFTMAOQITDINARAGSSATL
DARTQVRYVSGNSQVRFADIONMEVNTVPDQAKVIVNSRTGSVVMNRVTLDSCA
VAGNLSVTYVQNLVNPQNPTRFGGQIVTVPTQIDLRQSGSLQSVSSANLNSVY

```

RAINAIGATPMDLSILQSMQAGCLRAKLEII"
complement(5597. .6295)
/gene="figh"
/locus_tag="t1740"
complement(5597. .6295)
/gene="figh"
/locus_tag="t1740"
/note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MOKRALHVPYMAALMTLTCGAPITPAKPLVQAGATTAPQIPGV
PVANSIRFSAQPIVYQPLFEDRPRNIGTITVIGENVSASSKSSANRDKT
SPGFPIVPRYLOGFLGNSRADMEASFNKGASNAISNTFSGLTITVIVDLANGN
LHVGKQIAINQGTETFIKSVNPRITSGNSVPTQVADARIEYVNGYINBAON
MGLQRFPLNLSPM"
complement(6350. .7132)
/gene="figg"
/locus_tag="t1741"
complement(6350. .7132)
/gene="figg"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein figg"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISSIMIKTGLDQAQTMVDYIANLVSTNGEKGQAVEDL
LYQIRQPSAQSSQETLPSGLQIGVAPVATERLHQGNLSQTNNSKDAVAKQGF
FQVMLPDGTSATYRDSQVNDQGLVAGQVQVQVPAITIPANALITIGRDVSVT
OQGAAPVQVQGLNLTFTMNDRGLEISGENLYETQSSGAPNESTPGNLGAGLYQGY
VETSNVAEELNMIIQVRAEINSKAVSTTDQMLQKLTQL"
complement(7146. .7901)
/gene="figf"
/locus_tag="t1742"
complement(7146. .7901)
/gene="figf"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein figf"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHAIYTMGAASQTLNQAVTASNLAASTPGFRALNLRV
PVDGLATRLVLTASTPGADMTQGLDYTSRPLDVALQQDGMVLAQADAGEYTRN
GNIQVPTGOLTIQHPVIGEGPTIVPEGSPITTAAGQITSAALPGDPNIVAVGR
LKIYVAEENVEORSDGFLRLTAQAERGAETLADPSIRIMSGVLBSNVVPEAMT
DMIAARPEMQMKVITSYDENEGRANLLSMS"
complement(7922. .9133)
/gene="fige"
/locus_tag="t1743"
complement(7922. .9133)
/gene="fige"
/locus_tag="t1743"
/note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein fige"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFSQAVSGLMAATNLDVIGNNANSATTYFKSGTASPADMA
GSKVLGVKAVAGITDFTDGTNTNGRGLDAVISONGFRLVDSNGSVFYSRNGQFKL

Query Match 87.4%; Score 398.4; DB 1; Length 301983;
Best Local Similarity 92.1%; Pred. No. 5.3e-101;

Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGCGAGTCTTGCT 60
DB 37310 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGCGAGTCTTGCT 37251
QY 61 GGGCGGTTCACATATGGGGGGGGGGGGTATCATTAACGGGGGGGAATGTTCCGGC 120
DB 37250 GGGCGGTTCACATATGGGGGGGGGGGGTATCATTAACGGGGGGGAATGTTCCGGC 37191
QY 121 CCGGACTATGATCAGCTGTTTACCCGTGTGTGTAACCAATGAGCAGTCACTGCA 180
DB 37190 CCGGACTATGATCAGCTGTTTACCCGTGTGTGTAACCAATGAGCAGTCACTGCA 37131
QY 181 AGCGATGCCCGTAAATCTGAAAACGACATTAACCGAGCGGTTATGTAACGGCCCGAT 240
DB 37130 AGCGATGCCCGTAAATCTGAAAACGACATTAACCGAGCGGTTATGTAACGGCCCGAT 37071
QY 241 GTAGGCCAGGGTGGCGGATATATGATCTATGTAACGACTGCAAGATGTTTCAAAATAT 300
DB 37070 GTAGGCCAGGGTGGCGGATATATGATCTATGTAACGACTGCAAGATGTTTCAAAATAT 37011
QY 301 GCCACCATCGACAGTGGAGCGCTAAATCTCCATATTACTGTGCGCAATACCGCGGT 360
DB 37010 GCCACCATCGACAGTGGAGCGCTAAATCTCCATATTACTGTGCGCAATACCGCGGT 36951
QY 361 AATAACGCCGCGCTGTTAATCAGACCGCATCTGATTCGACGTAATGTCGTGCAAGTT 420
DB 36950 AATAACGCCGCGCTGTTAATCAGACCGCATCTGATTCGACGTAATGTCGTGCAAGTT 36891
QY 421 GGTTCGCAACACGCGCAGCGCTTACGATTTAA 456
DB 36890 GGTTCGCAACACGCGCAGCGCTTACGATTTAA 36855
RESULT 7
STAGFBA
LOCUS STAGFBA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AJ000514
VERSION AJ000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS Sukupolvi S., Lorenz R.G., Gordon J.I., Bian Z., Pfeiffer J.D.,
Normark S.J. and Rhen M.
TITLE Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE 2 (bases 1 to 1048)
AUTHORS Sukupolvi S.
TITLE Direct Submision
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kilnamyllykatu, 20520, FINLAND
FEATURES
source
1. 1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14. .553
/gene="agfB"
14. .553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="AgfB protein"
/protein_id="CAA04150.1"

```

/db_xref="GI:2275120"
/db_xref="SPTREMBL:O33801"
/translation="MKNLFLFMTLLTIGEPATATNYDLARSEYNFAVNELSKSFNO
AAILGQVGTDSARVRHSGSLSLVISOERGNNRAVDQAGNPNFATTCOTNADA
SISOAVGNSSAISAIAIIOKSGNKNANITGYGTAKTAVVVOKOSHAIQANITGYGTQ
KTAVVVOKOSHAIKRVTOR"
593..1048
/gene="agfa"
/cds
593..1048
/gene="agfa"
/codon_start=1
/translation="agfa"
/product="Agfa protein"
/protein_id="CAA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:O33802"
/translation="MKLLKVAFAIIVSGSAVAVVPQWGGGNGGNSGPDST
LSIYQVGSNAAALAJSDARKSETTITQSGYNGADVGGADNSTIETLTONPFRNAT
IDWNNKSDITVQYQYGNNAALVNOTASDSSVMVQVFGNNAIPANQYN"

```

Query Match 86.1%; Score 392.6; DB 1; Length 1048;
 Best Local Similarity 91.4%; Pred. No. 1.6e-99;
 Matches 416; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGATTGCGCAATGCTGATGTTCTGGCAGTCTCTGGCT 60
DB 593 ATGAACTTTTAAAGTGGCAGATTGCGCAATGCTGATGTTCTGGCAGTCTCTGGCT 652
QY 61 GGGGTGTTCCAAATGGGCGCGCGGATCATATACGGCGCGCAATAGTTCCGCG 120
DB 653 GGGGTGTTCCAAATGGGCGCGCGGATCATATACGGCGCGCAATAGTTCCGCG 712
QY 121 CCGGACTATGATCAGTGTGTTACCGGTGTTACCATGAAATGGCAATCATCATGCA 180
DB 713 CCGGACTTCCAGTGTGATGATTTATCAGTACGGTTCGCTAACGCTGCTCTGCA 772
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTACCCAGAGCGGTTATGTAACGGCGCAT 240
DB 773 AGCGATGCCGTTAAATCTGAAACGACATTACCCAGAGCGGTTATGTAACGGCGCAT 832
QY 241 GTAGCGCAGGTGCGGATATAGTACTATTAAGCACTCAAGATGTTTCAAAATAT 300
DB 833 GTAGCGCAGGTGCGGATATAGTACTATTAAGCACTCAAGATGTTTCAAAATAT 892
QY 301 GCCACATGACCAATGTAACCTTAAACTCCGATTTATCTGTCGTCATACGGCGGT 360
DB 893 GCCACATGACCAATGTAACCTTAAACTCCGATTTATCTGTCGTCATACGGCGGT 952
QY 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCACCGTTAATGTCGTCAGTT 420
DB 953 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCACCGTTAATGTCGTCAGTT 1012
QY 421 GGTTTGGCAACACGCGCATACCAAGTATTA 455
DB 1013 GGTTTGGCAACACGCGCATACCAAGTATTA 1047

```

RESULT 8
 LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
 DEFINITION Sequence 56 from patent US 5635617.
 ACCESSION 144908
 VERSION 144908.1 GI:2469621
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 361)
 AUTHORS Doran, J. L., Kay, W. W., Collinson, S. Karan, and Clouthier, S. C.
 TITLE Methods and compositions comprising the agfa gene for detection of
 Salmonella
 JOURNAL Patent: US 5635617-A 56 03-JUN-1997;

FEATURES
 source Location/Qualifiers
 1..361
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN

Query Match 61.9%; Score 282.2; DB 6; Length 361;
 Best Local Similarity 90.1%; Pred. No. 1.9e-68;
 Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 64 GTCTTCCAAATGAGGCGGCGGATTAATCATACGCGCGGCAATAGTCCGCCG 123
DB 1 GTCTTCCAAATGAGGCGGCGGATTAATCATACGCGCGGCAATAGTCCGCCG 60
QY 124 GACATATGATCAGCTGTTTACCGGTGTTTACCATGAAATGGCACATGCACTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTATTCAGTACGGTCCGCTAACGCTCGCTGTCGAAAGC 120
QY 184 GATCCCGTAAATCTGAAACGACCATTAACGAGGCGTTATGTAACGCGCCGATGTA 243
DB 121 GATCCCGTAAATCTGAAACGACCATTAACGAGGCGTTATGTAACGCGCCGATGTA 180
QY 244 GCGCAGGTGCGGATTAATGATTAATGAACTGACTCAGAAATGTTTCAAAATATGCC 303
DB 181 GCGCAGGTGCGGATTAATGATTAATGAACTGACTCAGAAATGTTTCAAAATATGCC 240
QY 304 ACCATGACACAGTGAACGCTAAACCTCGATTAATCTGTGGCCATACGCGGATAT 363
DB 241 ACCATGACACAGTGAACGCTAAACCTCGATTAATCTGTGGCCATACGCGGATAT 300
QY 364 AACCGCGCGGTGTTATGACACCGCATCTGATTC 398
DB 301 AACCGCGCGGTGTTATGACACCGCATCTGATTC 335

```

RESULT 9
 LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
 DEFINITION Citrobacter sp. Fec2 csbB gene, csbA gene and csbD gene.
 ACCESSION AJ515700
 VERSION AJ515700.1 GI:31790491
 KEYWORDS csbA gene; csbB gene; csbD gene; curlin-csbA protein; nucleation
 component of curlin monomers; regulatory protein.
 SOURCE Citrobacter sp. Fec2
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Citrobacter.
 1
 Zogaj, X., Bokranz, W., Nimtz, M. and Romling, U.
 Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
 Infect. Immun. 72 (7), 4151-4158 (2003)
 2 (bases 1 to 2889)
 Romling, U.
 Direct Submission
 Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
 Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES
 source Location/Qualifiers
 1..2889
 /organism="Citrobacter sp. Fec2"
 /mol_type="genomic DNA"
 /isolate="Fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:213763"
 /country="Germany"
 /complement(212..862)
 /gene="csbD"
 /complement(212..862)
 /gene="csbD"
 /function="necessary for transcription of the csbA
 operon"
 /codon_start=1

```

/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MFNEVHSIHGHTLLITRPSLOATALLQHLKQSALSGKLNIO
RSLDIISSCIVLDMEADKLIHYMDNLSRKNNLTLLNTPDDYPYDIENMP
HNGVFYVADDERVNVQGLRGECFSQKLASYLITHSGNRYNSSESLTHRE
KEILNKRIGASNIERSLFISENTVKTHLVNLFKIAVNRTOAVSANDNLNR"
1623..2078
/gene="csgb"
1623..2078
/gene="csgb"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKNKLFPMLTVLGPJASATSYDLAHEYNFAVNELSKSPN
QALIGOVYNNASAKRQSGSKLSVSGSGSNRAKQDSGAPAYIAQSGHSDA
SISOSNGNTAMIIQKSGSNKANITGYGKTAVVVQRQSMARIVTOR"
2119..2571
/gene="csga"
2119..2571
/gene="csga"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-csga protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLOVAFPAIVVSGSALAGVVPQWGGGGGGSSGSPSTL
SIYSGVNNALALQSDARKSPDTTHONGFNGAGVGGSDNSTIDLRNGKRNATT
DQNKNSDITVSYGHNALVNOTASDSVTVHVGFGNNTANQ?"

```

gene

CDS

Query Match 61.4%; Score 280; DB 1; Length 2889;
 Best Local Similarity 77.4%; Pred. No. 9e-68;
 Matches 353; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

```

OY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTCTGTGCT 60
DB 2119 ATGAACTTTTCAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTCTGTGCT 2178
OY 61 GCGCGCTTCCAGCATTTGGGCGCGCGGTATCATTAACGGGCGGCATAGTCCGGC 120
DB 2179 GGTTCGTGTCGCAATGGGGCGCGCGG--TGCGGGCGGGCGGAGCAGCTCCGCGC 2235
OY 121 CCGGACTATGATCAGCTGGTTACCCGTGTGTTTACCATGAATGAGCAATGCAATGCA 180
DB 2236 CCGGAAATGAGCCCTTATGATTCAGTCAGAGATCAATACCGCGCTTGCGTGC 2295
OY 181 AGCGATGCCGTTAAATCTGAAAGCAATTCACAGAGCGGTTATGTAACGGCGCCGAT 240
DB 2296 AGCGAGCTCGTAATCTGATACGACCAATTCAGAAATGCTTTGGTAACGGCGCAG 2355
OY 241 GTAAGCCAGGGTGGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
DB 2356 GTGGGCGAGGGTCCAGATTAACGACCAATTCGATCTCAAAAGCGGCTTCAAAAAC 2415
OY 301 GCGACATGACCACTGGAACGCTAAAACTCCGATATTACTGTGGGCCAATACGGCGGT 360
DB 2416 GCGACATGATCACTGGAACGCGCAAAATTCGACATTTACTGTGAGCCAGTATGTGA 2475
OY 361 AATAAGCCGCGCTGTTAATGACCGCATGTAATTCAGCGTAAATGTTGCGTCAAGTT 420
DB 2476 CATAAGCCGCGACTGTAACCAAGACTGCGTCCGATTCAGAGGTTCTGTGATCAGGT 2535
OY 421 GGTTTGGCAACAAGCGCAAGCTTAACAGATTAA 456
DB 2536 GGTTTGGCAACAAGCGCAAGCTTAACAGATTAA 2571

```

RESULT 10
 LOCUS AF515701 2920 bp DNA linear BCT 24-JUN-2003
 DEFINITION Citrobacter freundii csgB gene, csgA gene and csgD gene.
 ACCESSION AF515701
 VERSION AJ515701.1 GI:31790495
 KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.
 SOURCE
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
 1
 Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
 Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract Infect. Immun. 72 (7), 4151-4158 (2003)
 2 (bases 1 to 2920)
 Romling,U.
 Direct Submission
 Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 Location/Qualifiers
 1..2920
 /organism="Citrobacter freundii"
 /mol_type="genomic DNA"
 /isolate="fec4"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:546"
 /country="Germany"
 complement(211..861)
 /gene="csgD"
 complement(211..861)
 /gene="csgD"
 /function="necessary for transcription of the csgBA operon"
 /codon_start=1
 /transl_table=11
 /product="putative regulatory protein"
 /protein_id="CAD56673.1"
 /db_xref="GI:31790496"
 /translation="MFNEVHSIHGHTLLITRPSLOATALLQHLKQSALSGKLNIO
 RSLDIISSCIVLDMEADKLIHYMDNLSRKNNLTLLNTPDDYPYDIENMP
 HNGVFYVADDERVNVQGLRGECFSQKLASYLITHSGNRYNSSESLTHRE
 KEILNKRIGASNIERSLFISENTVKTHLVNLFKIAVNRTOAVSANDNLNR"
 1624..2082
 /gene="csgB"
 1624..2082
 /gene="csgB"
 /function="nucleator"
 /codon_start=1
 /transl_table=11
 /product="nucleation component of curlin monomers"
 /protein_id="CAD56674.1"
 /db_xref="GI:31790497"
 /translation="MKNKLFPMLTVLGPJASATSYDLAHEYNFAVNELSKSPN
 QALIGOVYNNASAKRQSGSKLSVSGSGSNRAKQDSGAPAYIAQSGHSDA
 SISOSNGNTAMIIQKSGSNKANITGYGKTAVVVQRQSMARIVTOR"
 2123..2572
 /gene="csgA"
 2123..2572
 /gene="csgA"
 /function="curli subunit"
 /codon_start=1
 /transl_table=11
 /product="curlin-csgA protein"
 /protein_id="CAD56675.1"
 /db_xref="GI:31790498"
 /translation="MKLOVAFPAIVVSGSALAGVVPQWGGGGGGSSGSPSTL
 SIYSGVNNALALQSDARKSPDTTHONGFNGAGVGGSDNSTIDLRNGKRNATT
 DQNKNSDITVSYGHNALVNOTASDSVTVHVGFGNNTANQ?"

ORIGIN

```

1
/moi_type="genomic DNA"
/strain="MC4100"
/db_xref="taxon:562"
3141. .4680
/organism="Escherichia coli"
/moi_type="genomic DNA"
/strain="W3110"
/db_xref="taxon:562"
<1. .>4680
/insertion_seq="IS2 (partial)"
complement [133. .966]
/gene="csqg"
complement [133. .966]
/gene="csqg"
/codon_start=1
/transl_table=11
/product="assembly /transport component in curli
production"
/protein_id="CA62277.1"
/db_xref="GI:1147559"
/db_xref="GOA:P52103"
/db_xref="SWISS-PROT:P52103"
/translacion="MORLELAUMLISCLTAPPEAKRPLTLMRAOSYKDLTHLP
FGKIKFVSVYNIODEGQKCPASNPASVPAVOSATAMLYTLKDSRPIPIERGL
NLINLKEKILRAAENGTVAINNRILPLOSILIANIVEGIIITYESNVSGYGAAP
IADYVDYQDQJAVNLARVYNVSTGELISVUMSKILSTYDAGVFRIDYORLEGG
VYVTSNEPMLCGLMAIEGVIFLINDGIDRLMDLONKAERONDILVKYRMSVPI
S"
complement [993. .1409]
/gene="csqg"
complement [993. .1409]
/gene="csqg"
/codon_start=1
/transl_table=11
/product="assembly /transport component in curli
production"
/protein_id="CA62278.1"
/db_xref="GI:1147560"
/db_xref="SWISS-PROT:P52104"
/translacion="MRVKAHVVLLMLISPLWAGTMTQFQPNPVFGGNPNNGALNS
AQONSXYKDSYNDPFEIETPSADLFTQAIQSOILGGLLSNTNGTGPGRMYNDYIV
DIARDDQLQNLVDRKTGOSTIQVSGLONNSTDP"
complement [1434. .1823]
/gene="csqg"
complement [1434. .1823]
/gene="csqg"
/codon_start=1
/transl_table=11
/product="assembly /transport component in curli
production"
/protein_id="CA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translacion="MRKYLKRWLVAAEFLPAAGNLHAVEVEPGLTIDHTVSSIGDHFY
RASDMSYSDYGNLTLYRMAFMSWTTITNOVDVIFQTEFLPIKROFEKTVVPAL
IQEELALNRQNALISGDLAHEF"
complement [1828. .2478]
/gene="csqg"
complement [1828. .2478]
/gene="csqg"
/notes="necessary for transcription of the csqgA operon"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translacion="WFEKHSIHGHTLLITKSSLOATFALQHLKOSIAITGKLNIQ
RSLDLSGSLITLIDMEADKLHYNODTLSRNNNIKILLNTPEPYRYDLENWP
HINQFFSMEDEQVANGLOADVLCGEFTQKLASYILITSGNRYNSTEALLTHRE
EELINKRIQASNNIEIARSIFISNTVKTHTLYNLFKKIAYKNRTOAVSMANDINLR"
complement [2626]

```

```

misc_feature      3140 /note="transcriptional start site for csdDEFG operon"
gene              3233..3688 /note="transcriptional start site for csdBA operon"
CDS              3233..3688 /gene="csdBA"
                  3233..3688 /gene="csdBA"
                  /codon_start=1
                  /transl_table=11
                  /product="nucleation component of curlin monomers"
                  /protein_id="CAA62281.1"
                  /db_xref="GI:1147563"
                  /db_xref="GOA:P39828"
                  /db_xref="SWISS-PROT:P39828"
                  /translation="MKKLLFMWLLITLGA PGIIRAAAGVLANSEYFNANSEKSPN
OAAITGAGTNNASQLRQGSKLAVVAOEGSSNPKIDQTDYINLAIYDQAGSANDA
SISQAGVNTAAITQKSGSNKANITQYGTQTAIVVORQSMARIVTOR"
                  3729..4184 /gene="csdA"
                  3729..4184 /gene="csdA"
                  /note="major subunit"
                  /codon_start=1
                  /transl_table=11
                  /product="curlin/CsdA protein"
                  /protein_id="CAA62282.1"
                  /db_xref="GI:1147564"
                  /db_xref="GOA:P28307"
                  /db_xref="SWISS-PROT:P28307"
                  /translation="MKLLKVAIAIAIVSGSALAGVDPYGGGNGHGGGNNSGPNSL
LNIYGGGNSALALQTDARNBDLITTHGGGNGADVGGSDSSIDLTORFGNSAT
LDVQKNSKSEMTVKQFGGGNGAAVDQTASNSVNTQYFGNNATAHQ"
                  3729..3788 /gene="csdA"
                  4243..4575 /gene="orfC"
                  4243..4575 /gene="orfC"
                  /gene="orfC"
                  /codon_start=1
                  /transl_table=11
                  /product="putative curli production protein"
                  /protein_id="CAA62283.1"
                  /db_xref="GI:1147565"
                  /db_xref="SWISS-PROT:P52107"
                  /translation="WNTLLILAASSQITPNTQGGDYTTITPEYTLTQSCICRYOIL
SURESQSQTQKRTKLSLPANOPTALTKLSLNTSPDRKIVVTVDGOSLHLSQO
WPPSEKSKS"
sig_peptide
gene
CDS

```

ORIGIN

Query Match 50.9%; Score 232; DB 1; Length 4680;

Best Local Similarity 69.3%; Pred. No. 3,1e-54; Matched 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGCGACGACATTCGAGCAATCGTATTCTGCGAGTCTGTGCT 60
    |||||
DB 3729 ATGAACCTTTTAAAGTGCGACGACATTCGAGCAATCGTATTCTGCGAGTCTGTGCT 3788
    |||||
QY 61 GCGCGTGTTCACAAATGGGCGCGCGCGGTAAATCATACGCGCGCGCAATAGTTCCGGC 120
    |||||
DB 3789 GGTGTGTTCCTCAGTACGCGCGCGCGGTAAATCATACGCGCGCGGTAAATAGCGGC 3848
    |||||
QY 121 CCGGACTATGATCAGTGGTTCACCGGTGTGTTATCCCATGAATAGGCATGACTGCA 180
    |||||
DB 3849 CCAAAATTCGAGCTGAACATTTACAGTACGCGGTGGGTTACTCTGCACTTCTCTGCA 3908
    |||||
QY 181 ACCGATGCCCGCTAAATCTGAAAGCAGCATATCCAGAGCGGTTATGTTAAAGCGCCCGAT 240
    |||||
DB 3909 ACTGATGCCCGCTAACTGACTGACTATTAATCCAGCAGCGGCGGTAAATGTCGCAAT 3968
    |||||
QY 241 GTAGGCGAGGGTGGGATATATGTAATTTGAACTGACTCAGAAATGTTTCAAAATTAAT 300
    |||||
DB 3969 GTTAGGTAGGGGCTCAGATGACGCTCAATCGATTCGACCAACGTCGTTCCGTAAACGC 4028
    |||||

```

```

QY 301 GCCACATCGACAGTGGAGCGTAAAACTCCGATATTAATGTCGCGCAATACGGCGGT 360
    |||||
DB 4029 GCTACTCTTGATCATGTGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGT 4088
    |||||
QY 361 AATAACGCCCGCGCTGTTAATGACACCGCATCTGATTCAGCGTAAATGATGTCGTGCT 420
    |||||
DB 4089 GGCACGCGGTCTGACGTTGACAGACGATCATTAATCTCTCCGTCACAGTACTCAGGTT 4148
    |||||
QY 421 GGTTTGGCACACGACCGACGCTAACCATTTTAA 456
    |||||
DB 4149 GGCCTTGTAAACACGACCGCTCATCATTA 4184
    |||||

```

RESULT 12

AE000205 10346 bp DNA linear BCT 01-DEC-2000
 Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE

Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
 The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)
 MEDLINE
 97426617
 9278503

REFERENCE

Blattner, F.R.
 Direct Submission
 Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecol@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
 3 (bases 1 to 10346)

REFERENCE

Blattner, F.R.
 Direct Submission
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecol@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
 4 (bases 1 to 10346)

REFERENCE

Plunkett, G. III.
 Direct Submission
 Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: markamber.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (http://cgsc.biology.yale.edu). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (http://www.genetics.wisc.edu). ** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications

and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'p' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

SOURCE

Location/Qualifiers
1. 10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
complement(<1. 48)
/note="153"

repeat_region

promoter

promoter

gene

CDS

CDS

/note="factor Sigma70; predicted +1 start at 1094689"
5. 32
/note="factor Sigma70; predicted +1 start at 1094716"
69. 392
/gene="b1028"
69. 392
/gene="b1028"
/function="orf; Unknown"
/note="0107; This 107 aa ORF is 27 pct identical (0 gaps) to 58 residues of an approx. 1888 aa protein PAS2_CANAL
SM: P43098"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74113.1"
/db_xref="GI:1787266"
/translation="MEDGYQAMRDVLAGDIASPHQVLMVATKSEVAPSIENUS
LHLETHVKTESLDLFTLMPPTVAFLEISNKPYLGNMLNDFAGVDQQRVHCRAFD
AVPRK"
275. 299
/gene="b1028"
/note="factor Sigma32; predicted +1 start at 1094983"
389. 1375
/gene="ycdu"
/note="synonym: b1029"
389. 1375
/gene="ycdu"
/function="orf; Unknown"
/note="0328; This 328 aa ORF is 19 pct identical (4 gaps) to 191 residues of an approx. 288 aa protein Y320_MYCSE
SM: P47562"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"
/translation="WIPDYLFIRFODKRNLIYVATGLIGPYWKNAGFTPSSEDI
GVSGILALVYNFIPDLKAYMAKCTKYNIDSPFKKONHKELELTPVAGPLS
LIMSAMSQVYOLPSLYALFISLIGVPLIFPMTFRTSVKQVALSPVAKYK
SLTRYVLSVCTISVNDLISPLRNSDSVTEQOMLTFKSIILALICGVVAIMLF
FLTRSKYAFGLFQIDLEIDLFSSENASLTFPAKPLMLLFTLLIVEMVITLVS
ATLVEWRIVFEAVYFLCYVPLIYFFYCRFLMHNDFMACDMYFRMGHNK"
1253. 1280
/gene="ycdu"
/note="central position to predicted promoter: -104"
/bound_moiety="Deor predicted site"
1253. 1280
/gene="ycdu"
/note="No predicted promoter"
/bound_moiety="Deor predicted site"
1337. 1365
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1096049"
1494. 1745
/gene="b1030"
1494. 1745

CDS

gene

CDS

/gene="b1030"
/function="orf; Unknown"
/note="083; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_RABIT
SM: P49065"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGTRCVQYTKKPYLSYELPFYKGGEGIDLSLSPCGQPAHC
VHGSLNLSPPVDPGRGLIPVCAIHEKARFVAL"
1513. 1926
/gene="ycdu"
/note="synonym: b1031"
1513. 1926
/gene="ycdu"
/function="putative structure; Not classified"
/note="0137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein R337_YEAST
SM: P05759"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKSDPYRTSSILMAVAGLTPRPARAGSPITAFYTCPT
GCRQSTPGRFSPPRCVQYTKKPYLSYELPFYKGGEGIDLSLSPCGQPAHC
SNMLSPVDPGRGLIPVCAIHEKARFVAL"
complement(2111. 2198)
/gene="sexX"
/note="synonym: b1032"
complement(2111. 2198)
/gene="sexX"
/product="rRNA-Sex"
/note="anticodon: GGA, CG Site No. 17878"
/function="RNA; tRNA"
2269. 2296
/note="factor Sigma70; predicted +1 start at 1096980"
2338. 2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355. 2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393. 3370
/gene="ycdu"
/note="synonym: b1033"
2393. 3370
/gene="ycdu"
/function="putative enzyme; Not classified"
/note="0325; This 325 aa ORF is 32 pct identical (2 gaps) to 177 residues of an approx. 312 aa protein YPRA_CORGL
SM: P45637"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="WRINFAQRFSGVSMIDILFHPFTDQMTLEALRKAIPQAVRAM
KSGNDSADVALVWHPPEMLAGDLRAVALGAVDSISLKIQHPHMLNPSVPLR
LEDTGSGQMOEYAVSQVLMFRFEDYRICOSSSHQDPEYHREDFTGIGAGVL
GSKYASIQTRFPLRCMSRTKSPGVQSGFAGSELSARLSQCRVILNLPPTPV
GIINOQLEKLPGAYLLNLARGVHYVEDLLAALSGKVKGLMVDVFNREPLPSP
LMQHPRTIIPHYAIIIRPAVAYISRTIAQLEKGRVCGQVDRARGI"
3278. 3308
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1097992"
3383. 3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396. 3423

protein_bind
protein_bind
promoter
gene
CDS

promoter
promoter
promoter

Query Match 50.9%; Score 232; DB 1; Length 10346;
Best Local Similarity 69.3%; Pred. No. 3.3e-54;
Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTTCTGGACGCTCTGGCT 60
 DB 8993 ATGAACCTTTTAAAGTGGCAGCAATTCGAGCAATCGTAGTTTCTGGACGCTCTGGCA 9052
 QY 61 GCGCTGTTTCCAAATGGGCGCGCGGTAATCATTAACGCGCGGCAATGTTCCGCG 120
 DB 9053 GGTGTTGTTTCCATGACGCGCGCGGTAATCATTAACGCGCGGTAATGTTCCGCG 9112
 QY 121 CCGGACTATGATCAGCTGTTAACCCTGTTTACCCATGAATGACATGCA 180
 DB 9113 CCAATTTCTGAGCTAAACATTACAGTAACGCTGCGGTAATGTTCTCTGCA 9172
 QY 181 AGCGATGCCCGTAATTCGAACGACATACCCAGAGCGGTTATGTTACGCGCCCAT 240
 DB 9173 ACTGATGCCCGTAATTCGACTGATTAATCCCGATGCGCGGTAATGTTCTCAAT 9232
 QY 241 GTAGCCAGGCTGCGGTAATGATTAATGTAATGTAATGTAATGTAATGTAAT 300
 DB 9233 GTTGTCTAGGCTCAGATGACAGCTCATTCGATCCAGCGGTTCCGTTAAGC 9292
 QY 301 GCCACCATGACAGCTGGAACGCTAAATCTCGATTAATGTTGCGCCATACGCGCT 360
 DB 9293 GCTACTCTGATCAGTGAACGCGCAAAATCTGAATGACGCTTAACAGTTCCGTT 9352
 QY 361 AATAAGCCGCGCTGTTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 420
 DB 9353 GCGAACGCTGCTGACGATGACGACGATGATTAATCTCTCGTCAAGTGAATGAT 9412
 QY 421 GATTGTCACACACGCGCTGATTAATGATTAATGATTAATGATTAATGATTAAT 456
 DB 9413 GCTTTGTTGATACACGCGCTGATTAATGATTAATGATTAATGATTAATGATTAAT 9448

FEATURES

source

Headed by:
 Name: Takashi Horuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishior@nibb.ac.jp
 Information operator:
 Name: Hirotsada Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmo@gtc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://bsw3.aist-nara.ac.jp.
 Location/Qualifiers
 1. 15047
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:83333"
 /map="23.7-24.0 min"
 /note="Nucleotide position 1097101-1112147 from the
 Kohara lambda miniset library."
 <1. 276
 /note="unlabeled protein product; ORF_ID:0230#5"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35822.1"
 /db_xref="GI:4062605"
 /translation="DLFSSSNALSTFPKPLMLRLFILVIEVMWITLVSLATLVE
 WRIVFAYFLICVYPCLIYFFCYRPLMNDPMWACDMYFPMGHFNK"
 1152. 2090
 /note="ORF_ID:0231#1
 similar to SwissProt Accession Number P45637"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35823.1"
 /db_xref="GI:4062606"
 /translation="MDIIFYHPPTDQWTEBALRKAIPQARVARKSGDNDSDAYLV
 WHPPVEMLAGDLKAVPALGAGVDSISKLQAEEMNPSPVPLRLSDTGSEMOEY
 AVSGLVWFRFRDRIQONSHPQPEYHREDFTIGAGVLSKVAQSLQTFR
 PLRWSRTRKSPVQSPAGRELSAFISCRVILNLPMPETVGIINQQLLEKLPD
 GAYLINLARGHVVEDLLALDSGKTKGMLDVFNRBPSPESPLMOPRVITTPHY
 AATRPFAVAYISRTIAQLKGRVCGQVDRAGY"
 2145. 2882
 /note="unlabeled protein product; ORF_ID:0231#2"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35824.1"
 /db_xref="GI:4062607"
 /translation="MYPVDLDMHTVASTHAYSTLSDYIAQKQKIKLFAITDHPDM
 EDAPHHMFTNMRIMPRVVDGILRGIEANIKVNDGIDCSGMPSPSLDIILGPHF
 PVFAPDKATVQMIATIASGVNHSHPNPKYEDIVKVAALAAKQVALLKPNNS
 SFLHSRSGEDNCEVAAAVRDAGWALGSDSHATFTWGFESCEKILDAVDPPPR

COMMENT

The Japan E. coli genome DNA sequencing project
 Group:
 The Japan E. coli genome DNA sequencing group

Members: (1995.4 - 1996.3)
 Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
 Horuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
 Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
 Kimura, S., Kitagawa, M., Kitagawa, M., Makino, K.,
 Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
 Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,
 Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
 Yamamoto, Y. and Yano, M.

Headed by:
 Name: Takashi Horuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishior@nibb.ac.jp
 Information operator:
 Name: Hirotsada Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmo@gtc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://bsw3.aist-nara.ac.jp.
 Location/Qualifiers
 1. 15047
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:83333"
 /map="23.7-24.0 min"
 /note="Nucleotide position 1097101-1112147 from the
 Kohara lambda miniset library."
 <1. 276
 /note="unlabeled protein product; ORF_ID:0230#5"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35822.1"
 /db_xref="GI:4062605"
 /translation="DLFSSSNALSTFPKPLMLRLFILVIEVMWITLVSLATLVE
 WRIVFAYFLICVYPCLIYFFCYRPLMNDPMWACDMYFPMGHFNK"
 1152. 2090
 /note="ORF_ID:0231#1
 similar to SwissProt Accession Number P45637"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35823.1"
 /db_xref="GI:4062606"
 /translation="MDIIFYHPPTDQWTEBALRKAIPQARVARKSGDNDSDAYLV
 WHPPVEMLAGDLKAVPALGAGVDSISKLQAEEMNPSPVPLRLSDTGSEMOEY
 AVSGLVWFRFRDRIQONSHPQPEYHREDFTIGAGVLSKVAQSLQTFR
 PLRWSRTRKSPVQSPAGRELSAFISCRVILNLPMPETVGIINQQLLEKLPD
 GAYLINLARGHVVEDLLALDSGKTKGMLDVFNRBPSPESPLMOPRVITTPHY
 AATRPFAVAYISRTIAQLKGRVCGQVDRAGY"
 2145. 2882
 /note="unlabeled protein product; ORF_ID:0231#2"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35824.1"
 /db_xref="GI:4062607"
 /translation="MYPVDLDMHTVASTHAYSTLSDYIAQKQKIKLFAITDHPDM
 EDAPHHMFTNMRIMPRVVDGILRGIEANIKVNDGIDCSGMPSPSLDIILGPHF
 PVFAPDKATVQMIATIASGVNHSHPNPKYEDIVKVAALAAKQVALLKPNNS
 SFLHSRSGEDNCEVAAAVRDAGWALGSDSHATFTWGFESCEKILDAVDPPPR

RESULT 13
 LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
 DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
 ACCESSION D90741 AB001340
 VERSION D90741.1 GI:1651509
 KEYWORDS Complete and shotgun sequencing; csq5; csqf; csqg; csqD; csqB;
 csqA; ycds; cys; nov; mdog.
 SOURCE Escherichia coli K12
 ORGANISM Escherichia coli K12
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1
 Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
 Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
 Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
 Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
 Nishio, Y., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K.,
 Wada, C., Yamamoto, Y., Yano, M. and Horuchi, T.
 A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map
 DNA Res. 3 (3), 137-155 (1996)
 JOURNAL MEDLINE 97061202
 PUBMED 8905232
 REFERENCE 2
 Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horuchi, T.,
 Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
 Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
 Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
 Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
 Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
 Yamamoto, Y. and Yano, M.
 The systematic sequencing of the Escherichia coli genome in Japan
 Unpublished
 JOURNAL 3 (bases 1 to 15047)
 REFERENCE 3
 Mori, H.
 Direct Submission
 Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
 and Technology, Res. & Edu. Center for Genetic Info., 8916-5

CDS

CDS

CDS

CDS
2906..3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein H11543"
/protein_id="BA35825.1"
/db_xref="GI:4062608"
/translation="MNEFSLICVLSLYRQPDLLVPLFTLREGKLANMPELQ
DELLTRLOKSCDMTOVSADYNALFIDECAPPEYRSAGWEAGTEAVERAFISERCMPL
ADTPADHIGITLLASMLDEQSTEDSEALETLFSEYLLPMCGAFLGKVEALATTPFW
RTMAPLTRDAISAMWDELDEDSSE"
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"
/codon_start=1
/transl_table=11
/protein_id="BA35826.1"
/db_xref="GI:4062609"
/translation="MCSISPMFAFSAHMGNNILSIATTTGILSGWVAVSLGL
SNAGFLGCTAYFACPOGGKGLAISATLLSGVWAMVTVGSALAPHLIELGYITG
IVAFMCIQAKOLLISFVPGTIGCATPAGGDKLVPLSLATGLITGYAMKSGLM
LAARSAKTHREDEIKKA"
complement(4117..4950)
/gene="cs9b"
complement(4117..4950)
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsgG precursor."
/protein_id="BAA35827.1"
/db_xref="GI:4062610"
/translation="MORPLIVAVMLSGCLTAPPKKAPRTLMRAOSYKDLTHLPA
PTGKIFVSYVNTODETGKFPYPASNESTAVQSAITVALDSDMFLEQGLQ
NUNERKTIIRAQENGVIAINRIPLQSLTANIMVESIIGYSNVAAGGVARYG
IGADTYQIDQIVANLRYVNVSTGELISVNTSKTILSYEVOAGVFPIIDORLLEE
VGVTSEPMCLMSAIEITGVIFLINDIDRGLMDLQKAKERQNDILVKRHMSVPE
S"
complement(4977..5393)
/gene="cs9f"
complement(4977..5393)
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsgF precursor."
/protein_id="BAA35828.1"
/db_xref="GI:4062611"
/translation="MRVHAVALMLISPLSMAGTWFQFRNPFSGNPNNGAFILNS
AQONSYKDPYNDPFIETPSALDNFTQALQGLILGGLSINTGKPRGMVNDYIV
DIARDQLOLVNDRKTGOTSTIQVSGLONNSTDP"
complement(5418..5807)
/gene="cs9g"
complement(5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsgE precursor."
/protein_id="BAA35829.1"
/db_xref="GI:4062612"
/translation="MKRYLWIVARFLPAAGLHAVEVEFQGLTDTVSGIDPFY
RASFQKESDYTNLTINERPSARKMSWITITVNDVITQTFPLPKRDEKTVFAL
IQTBELNRQIQALISTEDLHDEF"
complement(5812..6462)
/gene="cs9d"

gene
CDS

CDS
complement(5812..6462)
/gene="cs9d"
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=11
/product="Probable csgAB operon transcriptional regulatory
protein."
/protein_id="BAA35830.1"
/db_xref="GI:4062613"
/translation="MNEVSHIGHLTLITSSLOATLLOHLKQSLATITGKLNHQ
RSJDDISSGSIILLDMENADKLIHYWQDTLSKNNNITLILNTPEDYPYDIENMP
HINGVFSMEDQERVYVNGVQVLRGEYCFQKLSAYLITHSGVRYNSTESALATFRE
KEILNKLRIQASNNELARSFLISENTVKHLVNLFKLIAVKORIQAVSANNLNR"
7217..7672
/gene="cs9b"
7217..7672

gene
CDS

Query Match 50.3%; Score 232; DB 1; Length 15047;
Best Local Similarity 69.3%; Pred. No. 3,4e-54;
Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

Qy	1	ATGAACTTTAAAGAGGAGCATTCGAGCAATGTTGTCGCGCTGCTGCT	60
Db	7713	ATGAACTTTAAAGAGGAGCATTCGAGCAATGTTGTCGCGCTGCTGCA	7772
Qy	61	GCGCTGCTTCAATGAGGCGCGCGGTAATCAATACGCGCGCAATAGTCCGCG	120
Db	7773	GCTGTTGCTTCTCAGTACGCGCGCGCGGTAACACGCGTGTGCGGTAATAGCGGC	7832
Qy	121	CCGAGCTATGATGAGTGTGTTACCCGTGTTTACCATGAATGACACATGCTGCA	180
Db	7833	CCAAATTCAGTGAACATTTACAGTACGCTGCGGTAATCTGCACTTCTGCAA	7892
Qy	181	AGCGATGCCCTGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAGCGCGCAT	240
Db	7893	ACTGATGCCCGTATCTGACTGTTACTTACCCGACATGCGCGGTAATGTTCAAT	7952
Qy	241	GTAAGCCAGGTCGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT	300
Db	7953	GTTGTCAGGCGCTCAGATGACAGCTCAATGATGTAACCAACGTCGTTGCTGTAACGC	8012
Qy	301	GCCACATGACCAAGTGGACGCTAAATCCGATATTAATGTCGCGCATATAGCGCGCT	360
Db	8013	GCTACTCTTGAATCATGTGACGCGCAAAATTCGAAATGACGTTAAACAGTTGCGTGT	8072
Qy	361	AATTAACGCGCGCTGTTAATCAACGCGATCTGATTCAGCGTAAATGTCGTCAGGTT	420
Db	8073	GCGAACGCTGTCGACGTTGACCAAGCTGATCTTCTCGTCAACGTTACTCAGGTT	8132
Qy	421	GCTTTGGCAACACGCGACGCTTACCAATTTAA	456
Db	8133	GCTTTGTAACACGCGACGCTTACCAATTTAA	8168

RESULT 14
AX814811
LOCUS 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS
SOURCE
ORGANISM
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Bioerck, L., Olsen, A., Wikstrom, M. and Herwald, H.
Peptides
Patent: WO 03064446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
FEATURES
Source
1..456

/organism="Escherichia coli"
 /mol_type="unassigned DNA"
 /db_xref="taxon:562"
 1..456
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAE83541.1"
 /db_xref="GI:39104002"
 /translation="MKLIKVEAIIAIVPSSGALAGVDPYGGGNGHGGGNGNSGPNSE
 LNIYQGGNSALAQTDARNSDLITQHGNGADVQSGSDSIDLTQGFNSAT
 LDQWNGKNSMTVQKFGGNGAAVQDTASNSSVNTVQGFNNATAHQY"

ORIGIN

Query Match 50.5%; Score 230.4; DB 6; Length 456;
 Best Local Similarity 69.1%; Pred. No. 7.6e-54;
 Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGCAGTCTGGCT 60
 DB 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGCAGTCTGGCT 60
 QY 61 GCGCTGTTCCACAAATGGGCGCGCGGATATCAATACGCGCGGCAATAGTCCGAC 120
 DB 61 GGTGTGTTCTCAGTACGCGCGCGGATATCAATACGCGCGGCAATAGTCCGAC 120
 QY 121 CCGGACTATGATCAGCTGTTACCCGTGTTGTTACCAATGAAATGACATGCTGCA 180
 DB 121 CCAATTCGAGCTCAACATTTACGATACGATGCGGTTACTCTGCACTTGTCTGCA 180
 QY 181 AGCGATGCCCGGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAGCGCGCGAT 240
 DB 181 ACTGATGCCCGGTAATCTGAGCTTGAATTAACCGAGCGGTTATGTAAGCGCGAT 240
 QY 241 GTAGCCAGGCTGCGGATATGATCTATTAATGATCTGATCAATGATGTTTCAAAATAT 300
 DB 241 GTTGGTCAAGGCTCAGATGACAGCTCAATGATCTGATCAATGATGTTTCAAAAGC 300
 QY 301 GCCACATCGACGATGGAAGCTTAAATCCGATTTACTGTGGCCATACGCGCGT 360
 DB 301 GCTACTCTTGATCACTGGAACGCGCAAAATCTGAAATGACGCTTAAACAGTTCGCTGT 360
 QY 361 AATAAGCCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGATGTCAGGTT 420
 DB 361 GGCACGCTGCTGCACTTGCACGACGATCTGATCTCTCCGTCACGATGACTGAGTT 420
 QY 421 GGTTTGGCAACAAACGCGCAGCTTAACGATATTA 456
 DB 421 GGTCTTGTAAACAAACGCGCAGCTTAACGATATTA 456

RESULT 15
 ECOCGGA LOCUS 648 bp DNA linear BCT 13-JUL-1993
 DEFINITION Escherichia coli curlin subunit (csgA) gene, complete cds.
 VERSION L04979.1 GI:290424
 KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 648)
 Olesen, A., Arqvist, A., Hammar, M., Sukupolvi, S. and Normark, S.
 The RpoD sigma factor relieves H-NS-mediated transcriptional
 repression of csgA, the subunit gene of fibronectin-binding curli
 in Escherichia coli

JOURNAL Mol. Microbiol. 7 (4), 523-536 (1993)

MEDLINE 93211294
 PUBMED 8459772

COMMENT On Jun 11, 1993 this sequence version replaced gi:145630.
 Original source text: Escherichia coli (sub_strain W3110, strain
 K-12) (library: Kohara) DNA.

FEATURES

source

location/Qualifiers

1..648
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K-12"
 /sub_strain="W3110"
 /db_xref="taxon:562"
 /rdbase_1ib="Kohara"
 83..538
 /gene="csgA"
 83..538
 /gene="csgA"
 /note="precursor protein"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="curlin subunit"
 /protein_id="AAA23616.1"
 /db_xref="GI:290425"
 /translation="MKLIKVEAIIAIVPSSGALAGVDPYGGGNGHGGGNGNSGPNSE
 LNIYQGGNSALAQTDARNSDLITQHGNGADVQSGSDSIDLTQGFNSAT
 LDQWNGKNSMTVQKFGGNGAAVQDTASNSSVNTVQGFNNATAHQY"

CDS

sig_peptide
 /gene="csgA"
 /evidence=experimental
 143..535
 /gene="csgA"
 /product="curlin subunit"
 /evidence=experimental
 mat_peptide
 /gene="csgA"
 /evidence=experimental
 143..535
 /gene="csgA"
 /product="curlin subunit"
 /evidence=experimental

ORIGIN

Query Match 50.5%; Score 230.4; DB 1; Length 648;
 Best Local Similarity 69.1%; Pred. No. 7.6e-54;
 Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGCAGTCTGGCT 60
 DB 83 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGCAGTCTGGCT 60
 QY 61 GCGCTGTTCCACAAATGGGCGCGCGGATATCAATACGCGCGGCAATAGTCCGAC 120
 DB 143 GGTGTGTTCTCAGTACGCGCGCGGATATCAATACGCGCGGCAATAGTCCGAC 120
 QY 121 CCGGACTATGATCAGCTGTTACCCGTGTTGTTACCAATGAAATGACATGCTGCA 180
 DB 143 GGTGTGTTCTCAGTACGCGCGCGGATATCAATACGCGCGGCAATAGTCCGAC 180
 QY 121 CCGGACTATGATCAGCTGTTACCCGTGTTGTTACCAATGAAATGACATGCTGCA 180
 DB 203 CCAATTCGAGCTCAACATTTACGATACGATGCGGTTACTCTGCACTTGTCTGCA 262
 QY 181 AGCGATGCCCGGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAGCGCGCGAT 240
 DB 263 ACTGATGCCCGGTAATCTGAGCTTGAATTAACCGAGCGGTTATGTAAGCGCGAT 240
 QY 241 GTAGCCAGGCTGCGGATATGATCTATTAATGATCTGATCAATGATGTTTCAAAATAT 300
 DB 323 GTTGGTCAAGGCTCAGATGACAGCTCAATGATCTGATCAATGATGTTTCAAAAGC 382
 QY 301 GCCACATCGACGATGGAAGCTTAAATCCGATTTACTGTGGCCATACGCGCGT 360
 DB 383 GCTACTCTTGATCACTGGAACGCGCAAAATCTGAAATGACGCTTAAACAGTTCGCTGT 442
 QY 361 AATAAGCCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGATGTCAGGTT 420
 DB 443 GGCACGCTGCTGCACTTGCACGACGATCTGATCTCTCCGTCACGATGACTGAGTT 502
 QY 421 GGTTTGGCAACAAACGCGCAGCTTAACGATATTA 456
 DB 503 GGTCTTGTAAACAAACGCGCAGCTTAACGATATTA 456

Search completed: March 15, 2004, 22:50:13
 Job time : 1965.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcg.....ccacgctcaccagratraa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

```
1: geneeqn1980s:*
2: geneeqn1990s:*
3: geneeqn2000s:*
4: geneeqn2001s:*
5: geneeqn2001bs:*
6: geneeqn2002s:*
7: geneeqn2003as:*
8: geneeqn2003bs:*
9: geneeqn2003cs:*
10: geneeqn2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	3	AAC64625 Agfa::PT3
2	406.4	89.1	456	2	AA087467 Agfa segn
3	406.4	89.1	456	2	AA074142 Salmonell
4	406.4	89.1	456	3	AAC64617 Salmonell
5	363.2	79.6	456	3	AAC64626 Agfa::PT3
6	355.2	77.9	456	3	AAC64628 Agfa::PT3
7	355.2	77.9	456	3	AAC64622 Agfa::PT3
8	355.2	77.9	456	3	AAC64629 Agfa::PT3
9	355.2	77.9	456	3	AAC64623 Agfa::PT3
10	353.6	77.5	456	3	AAC64624 Agfa::PT3
11	352	77.2	456	3	AAC64630 Agfa::PT3
12	352	77.2	456	3	AAC64627 Agfa::PT3
13	348.8	76.5	456	3	AAC64631 Agfa::PT3
14	282.2	61.9	361	2	AA073066 Agfa segn
15	282.2	61.9	361	2	AA074141 Salmonell
16	230.4	50.5	456	3	AAC64619 Escherich
17	230.4	50.5	456	9	AAC64613 Agfa::PT3
18	218	47.8	646	2	AA062647 Fibronect
19	156	34.2	369	2	AA062646 FNB curli
20	51.2	11.2	100	7	ACD68807 E. coli K
21	50	11.0	78	3	AAC64609 Agfa (SEF
22	50	11.0	78	3	AAC64610 Agfa (SEF
23	48.2	10.6	78	9	ACF36151 E. coli C

24	48	10.5	48	3	AAC64621	AAC64621 Leishmani
25	43.2	9.5	48	3	AAC64616	AAC64616 S. enteri
26	42.4	9.3	100	7	ACD68808	ACD68808 E. coli K
27	41.2	9.0	78	3	AAC64606	AAC64606 SefA (SEF
28	40.8	8.9	100	7	ACD68809	ACD68809 E. coli K
29	40.6	8.9	78	3	AAC64605	AAC64605 SefA (SEF
30	39.4	8.6	47475	3	AAA81465	AAA81465 N. mening
31	39.4	8.6	110000	3	AAA81489_5	Continuation (6 of
32	39.4	8.6	349980	3	AA021612	Continuation (6 of
33	39.2	8.6	1083	5	AA076745	AA076745 DNA encod
34	37	8.1	14652	3	AAA81482	AAA81482 N. mening
35	37	8.1	110000	3	AAA81489_1	Continuation (2 of
36	37	8.1	349980	3	AA021611	Continuation (2 of
37	36.4	8.0	2610	7	ACA42270	ACA42270 Prokaryot
38	36	7.9	2667	7	ACA45315	ACA45315 Prokaryot
39	35.8	7.9	954	7	ACA25910	ACA25910 Prokaryot
40	35.6	7.8	2000	7	ADA71938	ADA71938 Rice gene
41	35.4	7.8	3300	6	AB067377	AB067377 Neisseria
42	35.4	7.8	3411	5	AA088526	AA088526 DNA encod
43	35.4	7.8	3412	5	AA089144	AA089144 DNA encod
44	34.8	7.6	2000	7	ADA71938	ADA71938 Rice gene
45	34.6	7.6	456	3	AAC64620	AAC64620 Escherich

ALIGNMENTS

RESULT 1
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX

DT 26-FEB-2001 (first entry)
XX

DE Agfa::PT3#4 DNA sequence SEQ ID NO:17.
XX

KW Salmonella; agfa; chromosomal gene replacement; fimbriin; epitope;
KM vaccine; immune response; immunogen; ds.
XX

OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX

PN WC200060102-A2.
XX

PD 12-OCT-2000.
XX

PF 05-APR-2000; 2000WO-CA000356.
XX

PI 05-APR-1999; 99US-0127888P.
XX

(UYVI-) UNIV VICTORIA.
XX

White AP, Doran JL, Collison SK, Kay WW;
XX

WPI; 2000-672631/65.
XX

P-PSDB; AAB36349.
XX

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
protein useful for eliciting immune response in animal.
XX

Disclosure; Page 136; 139pp; English.
XX

The present invention describes a recombinant agfa gene (I) where a
segment of the gene has been replaced by a segment of a foreign DNA
sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SfA1/TFP) nucleation depend
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on or segments of foreign amino acid polymer into the animal in conjunction with the host cell and introducing the bacteriophage host cell, from the host cell and introducing the useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrillar presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrillar subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrillae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.3e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
QY 61 GCGGTGTTCCACAAATGGGGGGGGGGGTAATATATACGGCGCGCAATAGTTCCGGC 120
DB 61 GCGGTGTTCCACAAATGGGGGGGGGGGTAATATATACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTATGATGACCTGCTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
DB 121 CCGGACTATGATGACCTGCTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
QY 181 AGCGATGCCCGTAATCTGAACGACATTTACCCAGAGCGGTTAAGGTAAACGGCGCGAT 240
DB 181 AGCGATGCCCGTAATCTGAACGACATTTACCCAGAGCGGTTAAGGTAAACGGCGCGAT 240
QY 241 GTAGGCGAGGGTCCGTAATATAGTACTTGAATGAGTCAAGATGTTTCAGAAATAT 300
DB 241 GTAGGCGAGGGTCCGTAATATAGTACTTGAATGAGTCAAGATGTTTCAGAAATAT 300
QY 301 GCCACCATGACAGTGAACGCTAAACCTCGATATTAATGTCGCGCAATACGGCGGT 360
DB 301 GCCACCATGACAGTGAACGCTAAACCTCGATATTAATGTCGCGCAATACGGCGGT 360
QY 361 AATAACGCGCGGCTGTTATATAGACCGCATCTGATTCAGCGTAATGTCGCGTCAAGTT 420
DB 361 AATAACGCGCGGCTGTTATATAGACCGCATCTGATTCAGCGTAATGTCGCGTCAAGTT 420
QY 421 GGTGTTGGCAACAGCGCAAGCTTAACCGATATTA 456
DB 421 GGTGTTGGCAACAGCGCAAGCTTAACCGATATTA 456

```

RESULT 2

AA087467 standard; DNA; 456 BP.

AA087467;

25-MAR-2003 (revised)
26-JUN-1995 (first entry)

Agfa sequence.

Salmonella; Agfa; vaccine; genetic immunization; ds.

Salmonella.

Key Location/Qualifiers
CDS 1..454
/tag= a
/note= "Agfa"

W09425598-A2.

10-NOV-1994.

26-APR-1994; 94WO-IB000207.

26-APR-1993; 93US-00054452.

(U)VI-1 UNIV VICTORIA INNOVATION & DEV CORP.

(KING/) KING J.

Kay WW, Collinson SK, Clouthier SC, Doran JL,

WPI, 1994-358275/44.

P-PSDB; MAR74625.

Eliciting an immune response to Salmonella - using attenuated Salmonella strains, vector constructs, or compans. conty. fibrillar type proteins.

Disclosure; Fig 7B; 95pp; English.

The DNA encodes the Salmonella Agfa protein. The DNA and isolated proteins are used in genetic immunization and vaccine compositions, respectively, to elicit an immune response to Salmonella in animals (e.g. food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN field.)

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 2; Length 456;
Best Local Similarity 93.2%; Pred. No. 9.9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
QY 61 GCGGTGTTCCACAAATGGGGGGGGGGGTAATATATACGGCGCGCAATAGTTCCGGC 120
DB 61 GCGGTGTTCCACAAATGGGGGGGGGGGTAATATATACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTATGATGACCTGCTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
DB 121 CCGGACTATGATGACCTGCTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
QY 181 AGCGATGCCCGTAATCTGAACGACATTTACCCAGAGCGGTTAAGGTAAACGGCGCGAT 240
DB 181 AGCGATGCCCGTAATCTGAACGACATTTACCCAGAGCGGTTAAGGTAAACGGCGCGAT 240
QY 241 GTAGGCGAGGGTCCGTAATATAGTACTTGAATGAGTCAAGATGTTTCAGAAATAT 300
DB 241 GTAGGCGAGGGTCCGTAATATAGTACTTGAATGAGTCAAGATGTTTCAGAAATAT 300
QY 301 GCCACCATGACAGTGAACGCTAAACCTCGATATTAATGTCGCGCAATACGGCGGT 360
DB 301 GCCACCATGACAGTGAACGCTAAACCTCGATATTAATGTCGCGCAATACGGCGGT 360
QY 361 AATAACGCGCGGCTGTTATATAGACCGCATCTGATTCAGCGTAATGTCGCGTCAAGTT 420
DB 361 AATAACGCGCGGCTGTTATATAGACCGCATCTGATTCAGCGTAATGTCGCGTCAAGTT 420
QY 421 GGTGTTGGCAACAGCGCAAGCTTAACCGATATTA 456
DB 421 GGTGTTGGCAACAGCGCAAGCTTAACCGATATTA 456

```

RESULT 3

AAT74142

ID	AA74142standard; DNA; 456 BP.
XX	
AC	AA74142;
XX	
DT	25-MAR-2003 (revised)
DT	29-SEP-1997 (first entry)
XX	
DE	Salmonella enteritidis 27655-3b agfa gene.
XX	
KX	Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX	
OS	Salmonella enteritidis.
XX	
FM	Key
FT	Location/Qualifiers
FT	1..456
CDS	
XX	
XX	US5635617-A.
XX	
PD	03-JUN-1997.
XX	
XX	26-APR-1994; 94US-00233788.
PF	
XX	26-APR-1993; 93US-00054452.
PR	
XX	(UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA	
XX	
PI	Collinson SK, Kay WW, Doran JL;
XX	
DR	WPI; 1997-309886/28.
XX	
XX	P-PSDB; AAM23570.
PT	
PT	Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
PT	enteropathogenic bacteria in the Enterobacteria family.
XX	
PS	Clam 1; Col 19-112; 85pp; English.
XX	
CC	The present sequence represents an isolated agfa gene derived from
CC	Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC	diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC	family Enterobacteria. It can also be used to provide proteins and
CC	antibodies which can be used for assays. The nucleic acid sequence can be
CC	used to provide probes or primers which can specifically hybridise to
CC	nucleic acid molecules from greater than 99% of Salmonella strains that
CC	are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC	from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC	correct PF field.)
XX	
XX	Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match	Similarity	89.1%	Score 406.4	DB 2	Length 456
Best Local	Similarity	93.2%	Pred. No. 9.9e-120		
Matches	Conservative	0	Mismatches 31	Indels	Gaps 0
QY	1 ATGAACCTTTTAAAGTGCGACACATTCCGACCAATCGTAGTTCTTGCGAGTCTCTGCT	60			
Db	1 ATGAACCTTTTAAAGTGCGACACATTCCGACCAATCGTAGTTCTTGCGAGTCTCTGCT	60			
QY	61 GCGCGTGTCCACATGGGGCGGGCGGGTATCATTAACGGGGCGGCGCAATGTCGGCG	120			
Db	61 GCGCGTGTCCACATGGGGCGGGCGGGTATCATTAACGGGGCGGCGCAATGTCGGCG	120			
QY	121 CCGGACTATGATCAGCTGGTTACCCGGTGTGTTATCCCATGAATGGCACATGCATGCAA	180			
Db	121 CCGGACTCAACGTTATAGCATTTATAGTACGGTTCCGCTAACGCTCGCTCTGCAA	180			
QY	181 AGCGATGCCCGTAATCTGAACGACCAATTACCAGAGCGGTTATGTTAGTAAACGGCGCCGAT	240			
Db	181 AGCGATGCCCGTAATCTGAACGACCAATTACCAGAGCGGTTATGTTAGTAAACGGCGCCGAT	240			
QY	241 GTAGGCCAGGGTGGCGATTAATGTACTATTGAACACTACAGAAATGTTTCAGAAATAAT	300			

[illegible]

RESULT 4
AAC64617
ID AAC64617 standard; DNA; 456 BP.
XX
XX
AC AAC64617;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis Agfa DNA sequence SEQ ID NO:1.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
MO2000060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000, 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
PI WPI; 2000-672631/65.
DR P-PSDB; AAB36341.
XX
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFA1/TF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 3; Length 456;
Best Local Similarity 93.2%; Pred. No. 9.9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGCGAGCATTTGCGAGCATGTGTTCTTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGCGAGCATTTGCGAGCATGTGTTCTTGGCAGTCTCTGCT 60
QY 61 GCGCTGTTCCACATGCGCGCGCGGCGGTAATCATAGCGCGCGCAATAGTCCGCG 120
DB 61 GCGCTGTTCCACATGCGCGCGCGGCGGTAATCATAGCGCGCGCAATAGTCCGCG 120
QY 121 CCGGACTATGATGAGTGTGTTACCGGTGTTGTTCCATGAAATGCAATGCACTGCAA 180
DB 121 CCGGACTATGAGTGTGTTACCGGTGTTGTTCCATGAAATGCAATGCACTGCAA 180
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTACCGAGCGGTATGTAACGCGCGCGAT 240
DB 181 AGCGATGCGCGTAAATCTGAAACGACATTACCGAGCGGTATGTAACGCGCGCGAT 240
QY 241 GTAGGCGAGGCGGTGATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGCGGTGATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACGAGTGAAGCGTAAATCTGCGATTTACTGTCGCAATGCGCGGT 360
DB 301 GCCACCATGACGAGTGAAGCGTAAATCTGCGATTTACTGTCGCAATGCGCGGT 360
QY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCGCGGTATGTCGCGAT 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCGCGGTATGTCGCGAT 420
QY 421 GATTGTCGCAACGCGCGCATGCTTAACCGATATTA 456
DB 421 GATTGTCGCAACGCGCGCATGCTTAACCGATATTA 456

```

RESULT 5

AA64626 standard; DNA; 456 BP.

AA64626;

26-FEB-2001 (first entry)

Agfa::PT3#5 DNA sequence SEQ ID NO:19.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

Escherichia coli.

Synthetic.

MO200060102-A2.

12-OCT-2000.

05-APR-2000; 2000WO-CA000356.

05-APR-1999; 99US-0127888P.

(UUVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay RW;

XX WPI: 2000-672631/65.
DR P-PSDB; AAB36350.
XX

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
protein useful for eliciting immune response in animal.
Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
segment of the gene has been replaced by a segment of a foreign DNA
(1) use of thin aggregative fimbriae (SEF17/TA) nucleation depended
assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Agfa*, *CsgA* and *Agfa*-homologue fimbria subunits, respectively; (2)
directing recombinant of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombinant of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC protein containing a replacement segment or segments of foreign amino
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC useful for the expression of recombinant with a carrier or diluent. (1) is
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 79.6%; Score 363.2; DB 3; Length 456;
Best Local Similarity 87.3%; Pred. No. 6.9e-106;
Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGCGAGCATTTGCGAGCATGTGTTCTTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGCGAGCATTTGCGAGCATGTGTTCTTGGCAGTCTCTGCT 60
QY 61 GCGCTGTTCCACATGCGCGCGCGGCGGTAATCATAGCGCGCGCAATAGTCCGCG 120
DB 61 GCGCTGTTCCACATGCGCGCGCGGCGGTAATCATAGCGCGCGCAATAGTCCGCG 120
QY 121 CCGGACTATGATGAGTGTGTTACCGGTGTTGTTCCATGAAATGCAATGCACTGCAA 180
DB 121 CCGGACTATGATGAGTGTGTTACCGGTGTTGTTCCATGAAATGCAATGCACTGCAA 180
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTACCGAGCGGTATGTAACGCGCGGT 240
DB 181 AGCGATGCGCGTAAATCTGAAACGACATTACCGAGCGGTATGTAACGCGCGGT 240
QY 241 GTAGGCGAGGCGGTGATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGCGGTGATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACGAGTGAAGCGTAAATCTGCGATTTACTGTCGCAATGCGCGGT 360
DB 301 GCCACCATGACGAGTGAAGCGTAAATCTGCGATTTACTGTCGCAATGCGCGGT 360
QY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCGCGGTATGTCGCGAT 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCGCGGTATGTCGCGAT 420
QY 421 GATTGTCGCAACGCGCGCATGCTTAACCGATATTA 456
DB 421 GATTGTCGCAACGCGCGCATGCTTAACCGATATTA 456

```


DB 421 GGTITGGCAACACGCCAGCTAACGATATTA 456

RESULT 6
AAC64628
ID AAC64628 standard; DNA; 456 BP.
AC AAC64628;
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#7 DNA sequence SEQ ID NO:23.
XX
XX Salmone11a: agfa; chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; de.
XX
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36352.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX
XX Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (Sef17/TRF) nucleation depended assembly system of strains of Salmone11a, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2) the directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmone11a, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbriin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 77.9%; Score 355.2; DB 3; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.5e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCAATTCGAGATTCTGAGTCTTGCGCT 60
DB 1 ATGAAACTTTTAAAGTGGCAGCAATTCGAGATTCTGAGTCTTGCGCT 60
QY GCGCGTCTTCCACAATGGGGCGGGCGGCTATCATACGCGCGCAATAGTCCGAC 120
DB 61 GCGCGTCTTCCACAATGGGGCGGGCGGCTATCATACGCGCGCAATAGTCCGAC 120
QY 121 CCGGACTATGATCAAGCTGCTTACCCGTTGTATACCATGAATGCACTGCGAA 180
DB 121 CCGGACTATGATGAGCATTTATCAGTACGCTTCCGTTACGCTGCTCTGCAA 180
QY 181 AGCGATGCGCGTAAATCGAAACGACATTCACGAGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATGCGCGTAAATCGAAACGACATTCACGAGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCGAGGTTGGGATTAATGATGATTAAGTCACTGCAATGTTTCAGAAATAT 300
DB 241 GTAGGCGAGGTTGGGATTAATGATGATTAAGTCACTGCAATGTTTCAGAAATAT 300
QY 301 GCCACCATCGACCAATGGAACGCTTAAATCTCGATATTAAGTCCCAATACGCGGT 360
DB 301 GCCACCATCGACCAATGGAACGCTTAAATCTCGATATTAAGTCCCAATACGCGGT 360
QY 361 AATTAAGCGCGGCTGTTAATCAGACCGCATCTGATTCAGAGTAAAGTCCGACGTT 420
DB 361 CATGAATGTCACATGCAATCAACGCGCATCTGATTCAGAGTAAAGTCCGACGTT 420
QY 421 GGTITGGCAACACGCCAGCTAACGATATTA 456
DB 421 GGTITGGCAACACGCCAGCTAACGATATTA 456

RESULT 7
AAC64622
ID AAC64622 standard; DNA; 456 BP.
AC AAC64622;
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#1 DNA sequence SEQ ID NO:11.
XX
XX Salmone11a: agfa; chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; de.
XX
XX Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36346.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX
XX Disclosure; Page 135; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of a recombinant
 CC protein containing a replacement segment or segments of a recombinant Agfa
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match Best Local Similarity 77.9%; Score 355.2; DB 3; Length 456;
 Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGAGTCTGCGCT 60
 DB 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGAGTCTGCGCT 60
 QY 61 GCGGCTTCGCAAAATGCGCGCGGCGGATATCAACGCGCGGCAATGTTCCGGC 120
 DB 61 GCGGCTTCGCAAAATGCGCGCGGCGGATATCAACGCGCGGCAATGTTCCGGC 120
 QY 121 CCGGACTATGATCAGCTGCTGTTACCCGCTGTTTACCATGAAATGCAATGCACTGCA 180
 DB 121 CCGGACTATGATCAGCTGCTGTTACCCGCTGTTTACCATGAAATGCAATGCACTGCA 180
 QY 181 AGCGATCCCGTAAATCTGAAACGACATTCACGAGCGGTTATGTAACGGCGCGAT 240
 DB 181 AGCGATCCCGTAAATCTGAAACGACATTCACGAGCGGTTATGTAACGGCGCGAT 240
 QY 241 GTAGGCCAGGCGCGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 DB 241 GTAGGCCAGGCGCGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 QY 301 GCCACCATGACCATGACATGAAAGCTTAAAGCTCGATATCTGCGGCAATGCGCGCT 360
 DB 301 GCCACCATGACCATGACATGAAAGCTTAAAGCTCGATATCTGCGGCAATGCGCGCT 360
 QY 361 AATTAACGCGCGGCTGTTATATCAACGCAATGATTCGAGTATGTTGCGTCAAGTT 420
 DB 361 AATTAACGCGCGGCTGTTATATCAACGCAATGATTCGAGTATGTTGCGTCAAGTT 420
 QY 421 GATTGGCAACAAGCGCGGCTAACCAATGATTA 456
 DB 421 GCAATGCAACAAGCGCGGCTAACCAATGATTA 456

RESULT 8
 ID AAC64629 standard; DNA; 456 BP.
 AC AAC64629;
 XX 26-FEB-2001 (first entry)
 DT
 XX

DE Agfa::PTJ#8 DNA sequence SEQ ID NO:25.
 XX Salmonella; agfa; chromosome1 gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 EN W0200060102-A2.
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay W;
 XX WPI: 2000-672631/65.
 DR P-PSDB; AAB36353.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX Disclosure; Page 138; 139p; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of a recombinant
 CC protein containing a replacement segment or segments of a recombinant Agfa
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match Best Local Similarity 77.9%; Score 355.2; DB 3; Length 456;
 Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGAGTCTGCGCT 60
 DB 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGAGTCTGCGCT 60
 QY 61 GCGGCTTCGCAAAATGCGCGCGGCGGATATCAACGCGCGGCAATGTTCCGGC 120
 DB 61 GCGGCTTCGCAAAATGCGCGCGGCGGATATCAACGCGCGGCAATGTTCCGGC 120
 QY 121 CCGGACTATGATCAGCTGCTGTTACCCGCTGTTTACCATGAAATGCAATGCACTGCA 180
 DB 121 CCGGACTATGATCAGCTGCTGTTACCCGCTGTTTACCATGAAATGCAATGCACTGCA 180

```
DB 121 CCGGACCTCAACGTGAGCATTTATCATGACGCTCCGTAACGCTGCGCTTATGATCAG 180
QY 181 AGCGATGCCCGCTAAATCTGAAACGACCATTTACCCAGAGCGGTTATGCTACGCCCGCAT 240
DB 181 CTGGTTACCCCGTGTGTTATCCCATGAAATGAGCATGAGGTTATGTAAACGCCCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAACTGACACTGAGATGTTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAACTGACACTGAGATGTTTTCAGAAATAT 300
QY 301 GCCACCATGACGACGTGAAACGCTAAACCTCCGATATTTACTGTCGCCCAATATACGCGCGT 360
DB 301 GCCACCATGACGACGTGAAACGCTAAACCTCCGATATTTACTGTCGCCCAATATACGCGCGT 360
QY 361 AATAACGCGCGCTGTTAATATGACACCGCATCTGATTCAGGTTATGCTGCTAGGTT 420
DB 361 AATAACGCGCGCTGTTAATATGACACCGCATCTGATTCAGGTTATGCTGCTAGGTT 420
QY 421 GGTTCGCAACAAACGCGCTAACCGATTTAA 456
DB 421 GGTTCGCAACAAACGCGCTAACCGATTTAA 456

RESULT 9
AAC64623
ID AAC64623 standard; DNA; 456 BP.
AC AAC64623;
XX
XX 26-FEB-2001 (first entry)
DE
XX Agfa::PT3#2 DNA sequence SEQ ID NO:13.
KM Salmone11a; agfa; chromosomal gene replacement; fimbriin; epitope;
KM vaccine; immune response; immunogen; ds.
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WM;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36347.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139p; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmone11a, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
```

```
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmone11a, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;
Query Match 77.9%; Score 355.2; DB 3; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.5e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 1 ATGAAACTTTTAAAGTGGACGACATTCGACCAATGTAAGTTTCTGGAGTCTTGCT 60
DB 1 ATGAAACTTTTAAAGTGGACGACATTCGACCAATGTAAGTTTCTGGAGTCTTGCT 60
QY 61 GCGCTGCTTCCAAATGAGGCGCGCGGCTAATCATTAACGCGCGCAATAGTTCCGCG 120
DB 61 GCGCTGCTTCCAAATGAGGCGCGCGGCTAATCATTAACGCGCGCAATAGTTCCGCG 120
QY 121 CCGGACTATGATCAGCTGTTTACCGGTGTTTACCCATGAAATAGGCATGACATGCA 180
DB 121 CCGGACTATGATGAGATTTATCATGACGTTCCGTAACGCTGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAAATCTGAAACGACATTCACCAAGCGGTTATGTAACGCGCGCAT 240
DB 181 AGCGATGCCCGTAAATCTGAAACGACATTCACCAAGCGGTTATGTAACGCGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAACTGACACTGAGATGTTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAACTGACACTGAGATGTTTTCAGAAATAT 300
QY 301 GCCACCATGACGACGTGAAACGCTAAACCTCCGATATTTACTGTCGCCCAATATACGCGCGT 360
DB 301 GCCACCATGACGACGTGAAACGCTAAACCTCCGATATTTACTGTCGCCCAATATACGCGCGT 360
QY 361 AATAACGCGCGCTGTTAATATGACACCGCATCTGATTCAGGTTATGCTGCTAGGTT 420
DB 361 CTGGTTACCCCGTGTGTTATCCCATGAAATGACATGCAACGTAATGCTGCTAGGTT 420
QY 421 GGTTCGCAACAAACGCGCTAACCGATTTAA 456
DB 421 GGTTCGCAACAAACGCGCTAACCGATTTAA 456

RESULT 10
AAC64624
ID AAC64624 standard; DNA; 456 BP.
AC AAC64624;
XX
XX 26-FEB-2001 (first entry)
DE
XX Agfa::PT3#3 DNA sequence SEQ ID NO:15.
KM Salmone11a; agfa; chromosomal gene replacement; fimbriin; epitope;
KM vaccine; immune response; immunogen; ds.
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO20060102-A2.
XX
```


CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred.No. 2.7e-102;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGACAGCATTCGACCAATCGATGTTCTGAGAGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGACAGCATTCGACCAATCGATGTTCTGAGAGTCTGCT 60
QY 61 GCGCTGTTCCACATGCGGCGCGCGGTATCTAAACGGCGGCAATGTTCCGCGC 120
DB 61 GCGCTGTTCCACATGCGGCGCGCGGTATCTAAACGGCGGCAATGTTCCGCGC 120
QY 121 CCGGACTATGATCAGCTGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCA 180
DB 121 CCGGACTATGATCAGCTGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCA 180
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTCACAGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCGCGTAAATCTGAAACGACATTCACAGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCCGAGGTCGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCCGAGGTCGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACGAGTGAACGCTTAAATCCGATTTATGTCGGCGCAATAGCGCGGT 360
DB 301 GCCACATCGACGAGTGAACGCTTAAATCCGATTTATGTCGGCGCAATAGCGCGGT 360
QY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAAATGTTGTCAGGTT 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAAATGTTGTCAGGTT 420
QY 421 GGTGTTGGCAACGCGCAGCGCTAACGATTAATTA 456
DB 421 GGTGTTGGCAACGCGCAGCGCTAACGATTAATTA 456

RESULT 12

AAC64627 standard; DNA; 456 BP.
AAC64627;
XX 26-FEB-2001 (first entry)
XX DT
XX Agfa::PT3#6 DNA sequence SEQ ID NO:21.
XX DE
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX KW vaccine; immune response; immunogen; ds.
XX OS
XX Salmonella enteritidis.
XX OS
XX Escherichia coli.
XX OS
XX Synthetic.
XX OS
XX WO20060102-A2.
XX PD
XX 12-OCT-2000.
XX PF
XX 05-APR-2000; 2000MO-CA000356.
XX PR
XX 05-APR-1999; 99US-0127888P.
XX XX
XX (UVI-) UNIV VICTORIA.
XX PI
XX White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
DR

DR P-PsDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP1/TFAP) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
Best Local Similarity 85.7%; Pred.No. 2.7e-102;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGACAGCATTCGACCAATCGATGTTCTGAGAGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGACAGCATTCGACCAATCGATGTTCTGAGAGTCTGCT 60
QY 61 GCGCTGTTCCACATGCGGCGCGCGGTATCTAAACGGCGGCAATGTTCCGCGC 120
DB 61 GCGCTGTTCCACATGCGGCGCGCGGTATCTAAACGGCGGCAATGTTCCGCGC 120
QY 121 CCGGACTATGATCAGCTGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCA 180
DB 121 CCGGACTATGATCAGCTGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCA 180
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTCACAGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCGCGTAAATCTGAAACGACATTCACAGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCCGAGGTCGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCCGAGGTCGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACGAGTGAACGCTTAAATCCGATTTATGTCGGCGCAATAGCGCGGT 360
DB 301 GCCACATCGACGAGTGAACGCTTAAATCCGATTTATGTCGGCGCAATAGCGCGGT 360
QY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAAATGTTGTCAGGTT 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAAATGTTGTCAGGTT 420
QY 421 GGTGTTGGCAACGCGCAGCGCTAACGATTAATTA 456
DB 421 GGTGTTGGCAACGCGCAGCGCTAACGATTAATTA 456

RESULT 13

AAC64631
ID AAC64631 standard; DNA; 456 BP.

XX AAC64631;

XX 26-FEB-2001 (first entry)

DE Agfa::PT#10 DNA sequence SEQ ID NO:29.

XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

XX vaccine; immune response; immunogen; ds.

XX Salmonella enteritidis.

XX Escherichia coli.

XX Synthetic.

XX MO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay MW;

XX WPI; 2000-672631/65.

XX P-PSDB; AAB36355.

XX Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/7AF) nucleation depended assembly system of strains of Salmonella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmonella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbria protein possesses both the immunogenicity and adhesion properties relevant for an efficient live immunogen, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match

Best Local Similarity 76.5%; Score 348.8; DB 3; Length 456;

Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGCGAGCATTCGCGAGCATCGTAGTTCTGCGAGTCTGTGCT 60

Db ||||| 1 ATGAACCTTTAAAGTGCGAGCATTCGCGAGCATCGTAGTTCTGCGAGTCTGTGCT 60

QY 61 GCGGTGTTCCACATGGGGGCGCGGGTATATATACCGCGGCAATAGTCCGCG 120

Db 61 GCGGTGTTCCACATGGGGGCGCGGGTATATATACCGCGGCAATAGTCCGCG 120

QY 121 CCGGACTATGATCAGCTGTACCCGTGTGTATACCAATATGACATGACTGCA 180

Db 121 CCGGACTATGATCAGCTGTACCCGTGTGTATACCAATATGACATGACTGCA 180

QY 181 AGCGATGCGGTAAATCTGAAAGACCATTAACCAAGCGGTTATGTAACGCGCGAT 240

Db 181 AGCGATGCGGTAAATCTGAAAGACCATTAACCAAGCGGTTATGTAACGCGCGAT 240

QY 241 GTAGGCCAGGTCGGATTAATAGTAACTAGTCAAGTGTGTTGCAATATAT 300

Db 241 GTAGGCCAGGTCGGATTAATAGTAACTAGTCAAGTGTGTTGCAATATAT 300

QY 301 GCCACCATGACCGAGTGAACGCTAAATCTCGATATTAATCTGTCGCAATACGCGGT 360

Db 301 GCCACCATGATAGCTGTATACCCGTGTGTATACCAATGACATGACGCGGT 360

QY 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCAGGTTATGTCGTCAGGTT 420

Db 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCAGGTTATGTCGTCAGGTT 420

QY 421 GGTTCGCAACAGCCGCGCTAACCATATTA 456

Db 421 GGTTCGCAACAGCCGCGCTAACCATATTA 456

RESULT 14

AAQ73066
ID AAQ73066 standard; DNA; 361 BP.

XX AAQ73066;

XX 27-AUG-2003 (revised)

XX 25-MAR-2003 (revised)

XX 26-JUN-1995 (first entry)

XX Agfa sequence.

XX Salmonella; Agfa; vaccine; genetic immunization; ds.

XX Salmonella enteritidis.

XX Key Location/Qualifiers

XX CDS 1..359

XX /*tag= a

XX /*note= "Agfa"

XX misc_feature 37..60

XX /*tag= d

XX /*note= "TAF5 primer (pair with TAF6)"

XX misc_feature 52..69

XX /*tag= b

XX /*note= "TAF3 primer (pair with TAF4)"

XX misc_feature complement(103..129)

XX /*tag= e

XX /*note= "TAF6 primer (pair with TAF5)"

XX complement(129..402)

XX misc_feature /tag= c

XX /note= "TAF4 primer (pair with TAF3)"

XX MO9425598-A2.

XX 10-NOV-1994.

XX 26-APR-1994; 94WO-IB000207.

XX 26-APR-1993; 93US-00054452.

XX

PA (UNIV-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
XX
PI Kay WW, Collinson SK, Clouthier SC, Doran JL,
XX WPI: 1994-358275/44.
XX P-PSDB; AAR62761.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compans. contg. fimbrial type proteins.
XX
XX Disclosure: Fig 7A; 95pp; English.
XX
CC The DNA encodes the Salmonella enteritidis 27655-3b TnpHox mutant strain
CC agfa gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to Salmonella in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 61.9%; Score 282.2; DB 2; Length 361;
Best Local Similarity 90.1%; Pred. No. 5.6e-80;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 64 GTCGTTCCACATATGGGGCGGGCGGCGGTATATCAATACCGCGGCAATAGTTCCGGCCCG 123
DB 1 GTCGTACACAGTGGGGCGGGCGGCGGTATATCAATACCGCGGCAATAGTTCCGGCCCG 60
XX
QY 124 GACTATGATCAGCTGTACCGGTGTGTTACCCATGAAATGGACATGCACTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATGATGACGTTCCGTTACCGTCCCTTGTCTGCAAGC 120
XX
QY 184 GATGCCCGTTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGCATGTA 243
DB 121 GATGCCCGTTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGCATGTA 180
XX
QY 244 GGGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 303
DB 181 GGGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 240
XX
QY 304 ACCATGACCAAGTGAACGCTAAACAACTCCGATATTACTGCGGCATATAGCGCGGTAT 363
DB 241 ACCATGACCAAGTGAACGCTAAACAACTCCGATATTACTGCGGCATATAGCGCGGTAT 300
XX
QY 364 AACGCGCGCTGTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCGCGCTGTGTTAATCAGACCGCATCTGATTC 335
XX

RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
AC AAT74141;
XX
DT 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
DE Salmonella enteritidis 27655-3b TnpHox mutant agfa gene fragment.
XX
KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
OS Salmonella enteritidis.
XX
FH Key Location/Qualifiers
FH CDS 1..360
FT /*tag= a
FT /label= agfa_gene_fragment
FT primer_bind 16..60
FT /*tag= b
FT /label= Primer_TAF5

FT primer_bind 52..69
FT /*tag= C
FT /label= Primer_TAF3
FT primer_bind complement(103..128)
FT /*tag= d
FT /label= Primer_TAF6
FT primer_bind complement(294..312)
FT /*tag= e
FT /label= Primer_TAF4
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UNIV-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX WPI: 1997-309886/28.
XX P-PSDB; AAW23569.
XX
PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
PT enteropathogenic bacteria of the Enterobacteria family.
XX
PS Claim 1; Col 107-110; 85pp; English.
XX
CC The present sequence represents an isolated agfa gene fragment derived
CC from Salmonella enteritidis 27655-3b TnpHox mutant strain. The nucleic
CC acid can be used to provide diagnostic assays for Salmonella and/or
CC enteropathogenic bacteria of the family Enterobacteria. It can also be
CC used to provide proteins and antibodies which can be used for assays. The
CC nucleic acid sequence can be used to provide probes or primers which can
CC specifically hybridise to nucleic acid molecules from greater than 99% of
CC Salmonella strains that are pathogenic to warm-blooded animals relative
CC to nucleic acid molecules from virtually all other microbial organisms.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 61.9%; Score 282.2; DB 2; Length 361;
Best Local Similarity 90.1%; Pred. No. 5.6e-80;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 64 GTCGTTCCACATATGGGGCGGGCGGCGGTATATCAATACCGCGGCAATAGTTCCGGCCCG 123
DB 1 GTCGTACACAGTGGGGCGGGCGGCGGTATATCAATACCGCGGCAATAGTTCCGGCCCG 60
XX
QY 124 GACTATGATCAGCTGTACCGGTGTGTTACCCATGAAATGGACATGCACTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATGATGACGTTCCGTTACCGTCCCTTGTCTGCAAGC 120
XX
QY 184 GATGCCCGTTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGCATGTA 243
DB 121 GATGCCCGTTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGCATGTA 180
XX
QY 244 GGGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 303
DB 181 GGGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 240
XX
QY 304 ACCATGACCAAGTGAACGCTAAACAACTCCGATATTACTGCGGCATATAGCGCGGTAT 363
DB 241 ACCATGACCAAGTGAACGCTAAACAACTCCGATATTACTGCGGCATATAGCGCGGTAT 300
XX
QY 364 AACGCGCGCTGTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCGCGCTGTGTTAATCAGACCGCATCTGATTC 335
XX

Search completed: March 15, 2004, 17:52:01

Thu Mar 18 12:28:00 2004

Job time : 246.684 secs

us-09-543-407-17.rng

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaacttcaaaagtcgc.....ccacgcctacacgtrataa 456

Scoring table:

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/6D.COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406.4	89.1	456	1	US-08-233-788A-58
2	282.2	61.9	361	1	US-08-233-788A-56
3	36.4	8.0	798	4	US-09-252-991A-13717
4	36.4	8.0	2748	4	US-09-252-991A-13830
5	36.4	8.0	3720	4	US-09-252-991A-13627
6	33.2	7.3	4403765	3	US-09-103-840A-2
7	33.2	7.3	4411529	4	US-09-103-840A-1
8	33.2	7.1	834	4	US-08-469-412A-8
9	32.4	7.1	332	3	US-09-021-715-8
10	32.4	6.9	2007	4	US-09-643-597-153
11	31.4	6.9	2007	4	US-09-480-884A-153
12	31.4	6.9	2007	4	US-09-542-615A-153
13	31.4	6.9	2007	4	US-09-606-421B-153
14	31.4	6.9	2007	4	US-09-221-107-153
15	31.4	6.9	2007	4	US-09-643-597-154
16	31.4	6.9	2148	4	US-09-480-884A-154
17	31.4	6.9	2148	4	US-09-542-615A-154
18	31.4	6.9	2148	4	US-09-606-421B-154
19	31.4	6.9	2148	4	US-09-221-107-154
20	31.2	6.8	1212	4	US-09-252-991A-7452
21	31.2	6.8	1425	4	US-09-252-991A-7188
22	31.2	6.8	1491	6	5466473-3
23	31.2	6.8	1692	4	US-09-252-991A-7513
24	31.2	6.8	2436	4	US-09-540-236-492
25	31.2	6.8	65792	4	US-09-596-002-31
26	30.8	6.8	2547	3	US-08-508-761B-1
27	30.8	6.8	2547	3	US-08-508-761B-1

28	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appl
29	30.4	6.7	552	4	US-09-252-991A-9800	Sequence 9800, Ap
30	30.4	6.7	894	4	US-09-434-354-1	Sequence 1, Appl
31	30.2	6.6	774	4	US-09-252-991A-11893	Sequence 11893, A
32	30.2	6.6	1083	4	US-09-252-991A-11930	Sequence 11930, A
33	30.2	6.6	3323	2	US-08-422-699A-10	Sequence 10, Appl
34	30.2	6.6	3323	2	US-08-422-706B-10	Sequence 10, Appl
35	29.8	6.5	1500	4	US-09-252-991A-6254	Sequence 6254, Ap
36	29.8	6.5	1695	4	US-09-252-991A-6044	Sequence 6044, Ap
37	29.8	6.5	1747	4	US-09-566-921-56	Sequence 66, Appl
38	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
39	29.4	6.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
42	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
43	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, Ap
44	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, Ap
45	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 3633617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELETYPE: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58

Query Match 89.1%; Score 406.4; DB 1; Length 456;
Best Local Similarity 93.2%; Pred. No. 3.1e-124;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGACAGATTCCGACCAATCGTAGTTCTGGAGTGTCTGGCT 60
Db 1 ATGAAACTTTTAAAGTGGACAGATTCCGACCAATCGTAGTTCTGGAGTGTCTGGCT 60
QY 61 GGCGTGTTCACCAATGGGGGGGGGGGGTAAATCAATAAGCGCGGCAATAGTTCCGAC 120
Db 61 GGCGTGTTCACCAATGGGGGGGGGGGGTAAATCAATAAGCGCGGCAATAGTTCCGAC 120
QY 121 CCGAGCTATGATCAGCTGGTTACCCGTTGTTGTTACCCATGAATGAGCATGACAGCA 180
Db 121 CCGAGCTATGATCAGCTGGTTACCCGTTGTTGTTACCCATGAATGAGCATGACAGCA 180
QY 181 AGCGATGCGCGTAAATCGAAGACCAATACCAAGCGGTTATGTTAAAGCGCGCAT 240
Db 181 AGCGATGCGCGTAAATCGAAGACCAATACCAAGCGGTTATGTTAAAGCGCGCAT 240
QY 241 GTAGGCGGAGGTGGGATTAATAGTACTATGAACTGACATGATGTTTCAAGAAATAT 300
Db 241 GTAGGCGGAGGTGGGATTAATAGTACTATGAACTGACATGATGTTTCAAGAAATAT 300
QY 301 GCCACCATTCACCGAGTGAACCGTAAATCTCGATATTAATGTCGCGCAATACGCGGT 360
Db 301 GCCACCATTCACCGAGTGAACCGTAAATCTCGATATTAATGTCGCGCAATACGCGGT 360
QY 361 AATACCGCGCGGTGTTATATCAGACCGCATCTGATTCAGAGGTATGTTGCGTCAAGTT 420
Db 361 AATACCGCGCGGTGTTATATCAGACCGCATCTGATTCAGAGGTATGTTGCGTCAAGTT 420
QY 421 GGTTTGGCAACAACCGCGCTAACCGATTAATCAAGTATTA 456
Db 421 GGTTTGGCAACAACCGCGCTAACCGATTAATCAAGTATTA 456

```

RESULT 2

US-08-233-788A-56
 Sequence 56, Application US/08233788A
 Patent No. 5635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, Karen S.
 APPLICANT: Clouthier, Sharon C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 TITLE OF INVENTION: OF SALMONELLA
 NUMBER OF SEQUENCES: 61
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043, 403C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEEDANBERRY
 INFORMATION FOR SEQ ID NO: 56:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 base pairs

TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..357
 US-08-233-788A-56

Query Match 61.9%; Score 282.2; DB 1; Length 361;
 Best Local Similarity 90.1%; Pred. No. 2,5e-83;
 Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 64 GTGCTTACCAATGGGGGGGGGGGGTAAATCAATAAGCGCGGCAATAGTTCCGAC 123
Db 1 GTGCTTACCAATGGGGGGGGGGGGTAAATCAATAAGCGCGGCAATAGTTCCGAC 123
QY 124 GACTATGATCAGCTGGTTACCCGTTGTTGTTACCCATGAATGAGCATGACATGCA 183
Db 61 GACTCAACGTGAGCATTTATAGTACGTTCCGTTACCGCTGCTGCTGCAAAAGC 120
QY 184 GATGCGCGTAAATCTGAAACGACATTACCAAGCGGTTATGTTAAAGCGCGCAT 243
Db 121 GATGCGCGTAAATCTGAAACGACATTACCAAGCGGTTATGTTAAAGCGCGCAT 240
QY 244 GGCGAGGTGGGATTAATAGTACTATGAACTGACATGATGTTTCAAGAAATAT 303
Db 181 GGCGAGGTGGGATTAATAGTACTATGAACTGACATGATGTTTCAAGAAATAT 300
QY 304 ACCATGACCATGGAACGCTAAATCTCGATATTAATGTCGCGCAATACGCGGT 363
Db 241 ACCATGACCATGGAACGCTAAATCTCGATATTAATGTCGCGCAATACGCGGT 360
QY 364 AACGCGCGGTGTTATATCAGACCGCATCTGATTC 398
Db 301 AACGCGCGGTGTTATATCAGACCGCATCTGATTC 335

```

RESULT 3

US-09-252-991A-13717/c
 Sequence 13717, Application US/09252991A
 Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 13717
 LENGTH: 798
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13717

Query Match 8.0%; Score 36.4; DB 4; Length 798;
 Best Local Similarity 47.4%; Pred. No. 0.04;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

```

QY 17 TGGAGATTCGACGAATCGTAGTTCTGGAGTGTCTGGCTGTTCCACAT 76
Db 270 TTGGCGGCTCGGCGAGCTGTGAAGGCGTGAACATGACAGGAGTGTTCAGTGT 211
QY 77 GGCGCGCGCGGTGTTATATCATAAGCGCGCGGCAATAGTTCCGCGGCAATAGTAC 136
Db 210 TCGGCGCGCGGTGTTGCGCGAGTGTGTCAGACAGTGCACAGGAGTGCAGTGTG 151
QY 137 TGGTTACCGCGTGTGTTTACCCATGAATGACATGACACTGCAAGCGATGCCGTAAAT 196

```

Db 150 TAGCCGCGCTGATGCTGCGACGCTGCGCGGACGTTGCTTGCATTAAGGGGGGAG 91
 QY 197 CTGAACGACCATTAACCCAGACCGGTTATGTTAAGCGCCCATATGAGC 246
 Db 90 GTGGCTTGGCCCTTGACGATGCGGAAAGAAACCGGCTTGACGTAGGC 41

RESULT 4 US-09-252-991A-13820/C

Sequence 13820, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 13820
 LENGTH: 2748
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13820

Query Match 8.0%; Score 36.4; DB 4; Length 2748;
 Best Local Similarity 47.4%; Pred. No. 0.074;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 17 TGGCAGCATTCGCGACCAATCGTAGTTCTGCGAGTCTCTGCGCGCTGCTTCCACAT 76
 Db 548 TTGGCGCGCTCGCGACGCTGCGAAGCGCGTGAACATCAGCAGGCTTGTGCTGAGTGT 489
 QY 77 GGGGGGGGGGGGGTATATTAACGGCGCGCAATAGTTCCGCGCCGACATATGATCAGC 136
 Db 488 TCGGCGCGCGTGTGTGCGCAGTTCGCGCTGCTCAGCAGCTCCACGAGGCGTGGCATGTTG 429
 QY 137 TGGTACCGCGTGTGTTACCCATGAAATGACATGACATGCAAGCGATGCCGTAAT 196
 Db 488 TAGCCGCGCTGATGCTGCGCGACGCTGCGCGGACGCTGCTTGTTCATTAAGGGGGGAG 369
 QY 197 CTGAACGACCATTAACCCAGACCGGTTATGTTAAGCGCGCCGATGAGC 246
 Db 368 GTGGCTTGGCCCTTGACGATGCGGAAAGAAACCGGCTTGACGTAGGC 319

RESULT 5 US-09-252-991A-13627

Sequence 13627, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 13627
 LENGTH: 3720
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13627

Query Match 8.0%; Score 36.4; DB 4; Length 3720;
 Best Local Similarity 47.4%; Pred. No. 0.087;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 17 TGGCAGCATTCGCGACCAATCGTAGTTCTGCGAGTCTCTGCGCGCTGCTTCCACAT 76
 Db 3080 TTGGCGCGCTCGCGACGCTGCGAAGCGCTGAACATCAGCAGGCTGTGCTTCACTGT 3139
 QY 77 GGGGGGGGGGGGGTATATTAACGGCGCGCAATAGTTCCGCGCCGACATATGATCAGC 136
 Db 3140 TCGGCGCGCGTGTGTGCGCAGTTCGCGCTGCTCAGCAGCTCCACGAGGCGTGGCATGTTG 3199
 QY 137 TGGTACCGCGTGTGTTACCCATGAAATGACATGACATGCAAGCGATGCCGTAAT 196
 Db 3200 TAGCCGCGCTGATGCTGCGCGACGCTGCGCGGACGCTTGTGCTTCAATTAAGGGGGGAG 3259
 QY 197 CTGAACGACCATTAACCCAGACCGGTTATGTTAAGCGCGCCGATGAGC 246
 Db 3260 GTGGCTTGGCCCTTGACGATGCGGAAAGAAACCGGCTTGACGTAGGC 3309

RESULT 6 US-09-103-840A-2

Sequence 2, Application US/09103840A
 Patent No. 6294328
 GENERAL INFORMATION:
 APPLICANT: FLEISCHMAN, Robert D.
 APPLICANT: WHITE, Owen R.
 APPLICANT: FRASER, Claire M.
 APPLICANT: VENTNER, John C.
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 TITLE OF INVENTION: TUBERCULOSIS
 FILE REFERENCE: 24366-20007.00
 CURRENT APPLICATION NUMBER: US/09/103,840A
 CURRENT FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 2
 LENGTH: 4403765
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 FEATURE:
 OTHER INFORMATION: CDC 1551
 OTHER INFORMATION: "n" bases at various positions throughout the sequence
 OTHER INFORMATION: represent a, t, c or g
 US-09-103-840A-2

Query Match 7.3%; Score 33.2; DB 3; Length 4403765;
 Best Local Similarity 44.6%; Pred. No. 31;
 Matches 131; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 77 GGGGGGGGGGGGGTATATTAACGGCGCGCAATAGTTCCGCGCCGACATATATCAGC 136
 Db 1093443 GCGGTGCGCGAGTGTGATGCGGAGCCCGGTGACGCTGTGTGCTGCGGCTCATCACCAGC 1093502
 QY 137 TGGTACCGCGTGTGTTACCCATGAAATGACATGACATGCAAGGATGCCGTAAT 196
 Db 1093503 ACCGTTGCTGTGAAATGCTACAGAACGCGCGGTGACGCGGCTTACCGCAC 1093562
 QY 197 CTGAACGACCATTAACCCAGACCGGTTATGTTAAGCGCGCCGATGAGCCAGGAGTGGC 256
 Db 1093563 CCGCTTGGACCGCTTGCAGATGCTATGCGGAGCGCGTGAACGCTGTGCTGCGGGGTTG 1093622
 QY 257 ATTAATGACTATTGAATCTGACGAATGTTTTCAGAAATTAATGCAACATGACACT 316
 Db 1093623 ATCAACGACACCGTGGCGCTGCCAATTTGCTGACGTAGACGTTGCCCGCGGTTGACC 1093682
 QY 317 GGAACGCTAAATACTCGGATATTACTGTGCGCAATACGCGGCTTAATGAGCGC 370
 Db 1093683 GCCACCGCGTGGGATGCTGTAAGCTGCGGCTGATGCTGATGAGCGCGC 1093736

RESULT 7

Query Match	7.3%;	Score 33.2;	DB 3;	Length 4411529;
Best Local Similarity	44.6%;	Pred. No. 31;		
Matches 131;	Conservative	0. Mismatch		

	1093444	1093445	1093446	1093447	1093448	1093449	1093450	1093451	1093452	1093453	1093454	1093455	1093456	1093457	1093458	1093459	1093460	1093461	1093462	1093463	1093464	1093465	1093466	1093467	1093468	1093469	1093470	1093471	1093472	1093473	1093474	1093475	1093476	1093477	1093478	1093479	1093480	1093481	1093482	1093483	1093484	1093485	1093486	1093487	1093488	1093489	1093490	1093491	1093492	1093493	1093494	1093495	1093496	1093497	1093498	1093499	1093500	1093501	1093502	1093503	1093504	1093505	1093506	1093507	1093508	1093509	1093510	1093511	1093512	1093513	1093514	1093515	1093516	1093517	1093518	1093519	1093520	1093521	1093522	1093523	1093524	1093525	1093526	1093527	1093528	1093529	1093530	1093531	1093532	1093533	1093534	1093535	1093536	1093537	1093538	1093539	1093540	1093541	1093542	1093543	1093544	1093545	1093546	1093547	1093548	1093549	1093550	1093551	1093552	1093553	1093554	1093555	1093556	1093557	1093558	1093559	1093560	1093561	1093562	1093563	1093564	1093565	1093566	1093567	1093568	1093569	1093570	1093571	1093572	1093573	1093574	1093575	1093576	1093577	1093578	1093579	1093580	1093581	1093582	1093583	1093584	1093585	1093586	1093587	1093588	1093589	1093590	1093591	1093592	1093593	1093594	1093595	1093596	1093597	1093598	1093599	1093600	1093601	1093602	1093603	1093604	1093605	1093606	1093607	1093608	1093609	1093610	1093611	1093612	1093613	1093614	1093615	1093616	1093617	1093618	1093619	1093620	1093621	1093622	1093623	1093624	1093625	1093626	1093627	1093628	1093629	1093630	1093631	1093632	1093633	1093634	1093635	1093636	1093637	1093638	1093639	1093640	1093641	1093642	1093643	1093644	1093645	1093646	1093647	1093648	1093649	1093650	1093651	1093652	1093653	1093654	1093655	1093656	1093657	1093658	1093659	1093660	1093661	1093662	1093663	1093664	1093665	1093666	1093667	1093668	1093669	1093670	1093671	1093672	1093673	1093674	1093675	1093676	1093677	1093678	1093679	1093680	1093681	1093682	1093683	1093684	1093685	1093686	1093687	1093688	1093689	1093690	1093691	1093692	1093693	1093694	1093695	1093696	1093697	1093698	1093699	1093700	1093701	1093702	1093703	1093704	1093705	1093706	1093707	1093708	1093709	1093710	1093711	1093712	1093713	1093714	1093715	1093716	1093717	1093718	1093719	1093720	1093721	1093722	1093723	1093724	1093725	1093726	1093727	1093728	1093729	1093730	1093731	1093732	1093733	1093734	1093735	1093736	1093737	1093738	1093739	1093740	1093741	1093742	1093743	1093744	1093745	1093746	1093747	1093748	1093749	1093750	1093751	1093752	1093753	1093754	1093755	1093756	1093757	1093758	1093759	1093760	1093761	1093762	1093763	1093764	1093765	1093766	1093767	1093768	1093769	1093770	1093771	1093772	1093773	1093774	1093775	1093776	1093777	1093778	1093779	1093780	1093781	1093782	1093783
--	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

```

RESULT 8
US-09-621-976-2574
/ Sequence 2574 Application US/09621976 .
/ Patent No 6639063
GENERAL INFORMATION:
/ APPLICANT: Dumas Malne Edwards, J.B.
/ APPLICANT: Jobert, S.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins
/ FILE REFERENCE: GENSET.054PR2
/ CURRENT APPLICATION NUMBER: US/09/621,976
/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 2574
/
/ LENGTH: 834
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 497..826
/ US-09-621-976-2574

```

Query Match 7.2%; Score 33; DB 4; Length 834;
Best Local Similarity 14.9%; Pred. NO. 0.54;
Matches 18; Conservative 64; Mismatches 39; Indels 0; Gaps 0

QY 79 GGGGGGGGGGTTAATCATATACGGCGGGCAATAGTCCGGCCCGAATTATGATCAGCTG 139
 Db 35 RGGGGGRRSISNNRRKNSYGARRCMSRRKAKCMWSMTGMSCTGRRMSWTKRRGMSIW 94
 QY 139 GTTACCCGCTGTTGTTACCCATGAATGCGACATGACTGCAAAAGGANTGCCGTTAATCT 189
 Db 95 STNNKRRGGSTYSYRMYRCYMSMMWMMCMRSCYRSBMMCMMSGRRSMCTGMSARCW 154
 QY 199 G 199
 Db 155 R 155

RESULT 9
US-08-469-412A-8

Patent No. 5856125
Application US/08469412A
GENERAL INFORMATION:
APPLICANT: Mavrothlaasitis, George J.
APPLICANT: Blair, Donald G.
APPLICANT: Fisher, Robert J.
APPLICANT: Beal Jr., Gregory J.
APPLICANT: Athanasios, Meropi A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS: 16
IND:

```

1 STREET: Two Embarcadero and Crew LLP
2 CITY: San Francisco
3 STATE: California
4 COUNTRY: USA
5 ZIP: 94111-3834
6
7 COMPUTER READABLE FORM:
8 MEDIUM TYPE: Floppy disk
9 COMPUTER: IBM PC compatible
10 OPERATING SYSTEM: PC-DOS/MS
11 SOFTWARE: PatentIn Release #1.0, Version #1.30
12 CURRENT APPLICATION DATA:
13 APPLICATION NUMBER: US/08/469,412A
14 FILING DATE: 05-JUN-1995
15 CLASSIFICATION: 435
16
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Garrett-Mackowski, Eugenia
19 REGISTRATION NUMBER: 37,330
20 REFERENCE/DOCKET NUMBER: 015280-229000
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (415) 576-0200
23 TELEFAX: (415) 576-0300
24
25 INFORMATION FOR SEQ ID NO: 8:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 332 base pairs
28 TYPE: nucleic acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 MOLECULE TYPE: DNA (genomic)
32 FEATURE:
33 NAME/KEY: -
34 LOCATION: 1..332
35 OTHER INFORMATION: /note= "human ERF gene promoter region"
36
37 US-08-469-412A-8

```

Query Match	7.1%;	Score	32.4;	DB	2;	Length	332;
Best Local Similarity	47.1%;	Pred.	No	0	53.		

Matches	99;	Conservative	0;	Mismatches	111;	Indels	0;	Gaps	0
---------	-----	--------------	----	------------	------	--------	----	------	---

Y
47 GCAGTGTCTATTGATGCCACCACTTTCCAACTTCCTCCTGGGCTGCCTG

100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918

120 GCGGGAGGGGCGGCGCAGTGTCTCCATGGCGACGCGCGGTGACCTGCGCGCGCGG 170

107 GCATAGTCCCGGCCGGAATTATCATCACCCTTTGGGCTTC

166

Db 180 GGGCGTGGGCTGCCGGGCGCCGAGTGCATATTAAACCGGAGGCGGCGGGGAGGG 239
Qy 167 CACATGCACTGCAAAAGCGATGCCCGTAATCTGAAACGACATTACCCAGACGGTTATG 226
Db 240 GAGAGCTCTTGAGAGCGGAGCGCGGTGTAGCGCGAGAGGCGCCGACGGGCGGGGAC 299
Qy 227 GTAACGGCGCCGATGTAGGCGCAGGCGG 256
Db 300 GGGACGGGCGACGAGGCGCGCGGAGCCG 329

RESULT 10
US-09-021-715-8
; Sequence 8, Application US/09021715
; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: Mavrochalaris, George J.
; Blair, Donald G.
; Fisher, Robert J.
; Beal Jr., Gregory J.
; Athanasiou, Meropi A.
; Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garett-Mackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; TELEPHONE: (415) 576-0200
; TELECOMMUNICATION INFORMATION:
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY:
; LOCATION: 1..332
; OTHER INFORMATION: /note="human ERF gene promoter region"
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-021-715-8

Query Match 7.1%; Score 32.4; DB 3; Length 332;
Best Local Similarity 47.1%; Pred. No. 0.53;
Matches 99; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 47 GCAGTCTCTGGCTGGCTGCTTCCACATAGGGGCGCGGTATCATTAACGGCGCG 106
Db 120 GCGGGAGGGGCGCGCGCATGTCTCCATGCGGAGCGCGGTACCTGCCGCGCGG 179
Qy 107 GCATAGTTCCGGCCCGGACATATGATCAAGCTGTTACCCGTGTGTTAACCATGAATGG 166
Db 180 GGGCGTGGGCTGCCGGGCGCCGAGTGCATATTAAACCGGAGGCGCGCGGAGGG 239

Qy 167 CACATGCACTGCAAAAGCGATGCCCGTAATCTGAAACGACATTACCCAGACGGTTATG 226
Db 240 GAGAGCTCTTGAGAGCGGAGCGCGGTGTAGCGCGAGAGGCGCCGACGGGCGGGGAC 299
Qy 227 GTAACGGCGCCGATGTAGGCGCAGGCGG 256
Db 300 GGGACGGGCGACGAGGCGCGCGGAGCCG 329

RESULT 11
US-09-643-597-153/c
; Sequence 153, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hoeken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 153
; LENGTH: 2007
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-643-597-153

Query Match 6.9%; Score 31.4; DB 4; Length 2007;
Best Local Similarity 46.5%; Pred. No. 2.8;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 119 GCCCGACTAGATGACGCTGTTACCGGTGTTTACCAATGATGACATGACATGC 178
Db 1337 GCTTGTATTAAGAACACATGTTCCCGAATAATCTTCAAAAGATTTTATTTT 1378
Qy 179 AAAGCATGCCCGTAATCTGAAACGACATTACCCAGAGCGGTTATGTAACGCGCG 238
Db 1377 AGTTAAAGTACTTAAAGTCACTGGGAATTTTGCAATGAGTGAAGAACTGAGCT 1318
Qy 239 ATGTAGCCAGGGTGGCGATATGTAATGTAAGTGAAGTCAAGATGCTTTCAGAAATA 298
Db 1317 GCTTAGGGGATTTACAGATTAGGCTTGAATTAATCAACAGAAATCCAGTATA 1258
Qy 299 ATGCCACATGACGAGTGAAGCTTAAATCTCGA 335
Db 1257 ATCTGCCCTCCATTAGTGTAGGAGAGAGCTCAA 1221

RESULT 12
US-09-480-884A-153/c
; Sequence 153, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Fan, Liqun
; APPLICANT: Hoeken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A

;; CURRENT FILING DATE: 2001-08-27
;; NUMBER OF SEQ ID NOS: 330
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 153
;; LENGTH: 2007
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-480-884A-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGAGTATGATCAGCTGTTACCGGTTGTTACCATGAATGCAATGCACCTGC 178
DB 1437 GCTTGTTCAATGAACACCATGTTCCCGAGAAAATCTCTACAAAGATGTTCAATTTT 1378
QY 179 AAGCGATGCGCCGTAATCTGAAGACCATTAACCCAGAGCGGTTATGTTAGCGCCG 238
DB 1377 AGTTAAAGTACTTAAAGTTCACTGGAAATTTTCAATGCAATGCAAGAAACTGGAGCT 1318
QY 239 ATGTAGCCAGGGTGGGATTAATAGTACTATTAAGTCACTGAGATGTTGAGAAATA 298
DB 1317 GCCTAGGGGAATTTTCAGATTAGGCTTGAATTACTCAACCCAGAGAAATCCAGATA 1258
QY 299 ATGCCACATCGACCAAGTGAAGCTTAAAACTCCGA 335
DB 1257 ATCTGCCCTCCCATTTAGTGTAGAGAGAGAGCTCCAA 1221

RESULT 13

US-09-542-615A-153/c
;; Sequence 153, Application US/09542615A
;; Patent No. 6518256
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Tonglong
;; APPLICANT: Fan, Liqun
;; APPLICANT: Kalos, Michael D.
;; APPLICANT: Bangur, Chaitanya S.
;; APPLICANT: Hosken, Nancy A.
;; APPLICANT: Fanger, Gary R.
;; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
;; FILE REFERENCE: 210121.455C8
;; CURRENT APPLICATION NUMBER: US/09/542,615A
;; NUMBER OF SEQ ID NOS: 350
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 153
;; LENGTH: 2007
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-542-615A-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGAGTATGATCAGCTGTTACCGGTTGTTACCATGAATGCAATGCACCTGC 178
DB 1437 GCTTGTTCAATGAACACCATGTTCCCGAGAAAATCTCTACAAAGATGTTCAATTTT 1378
QY 179 AAGCGATGCGCCGTAATCTGAAGACCATTAACCCAGAGCGGTTATGTTAGCGCCG 238
DB 1377 AGTTAAAGTACTTAAAGTTCACTGGAAATTTTCAATGCAATGCAAGAAACTGGAGCT 1318
QY 239 ATGTAGCCAGGGTGGGATTAATAGTACTATTAAGTCACTGAGATGTTGAGAAATA 298
DB 1317 GCCTAGGGGAATTTTCAGATTAGGCTTGAATTACTCAACCCAGAGAAATCCAGATA 1258
QY 299 ATGCCACATCGACCAAGTGAAGCTTAAAACTCCGA 335
DB 1257 ATCTGCCCTCCCATTTAGTGTAGAGAGAGAGCTCCAA 1221

RESULT 14
US-09-606-421B-153/c
;; Sequence 153, Application US/09606421B
;; Patent No. 6531315
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Tonglong
;; APPLICANT: Fan, Liqun
;; APPLICANT: Kalos, Michael D.
;; APPLICANT: Bangur, Chaitanya S.
;; APPLICANT: Hosken, Nancy
;; APPLICANT: Fanger, Gary R.
;; APPLICANT: Li, Samuel X.
;; APPLICANT: Wang, Aijun
;; APPLICANT: Skeiky, Yasir A.W.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;; FILE REFERENCE: 210121.455C9
;; CURRENT APPLICATION NUMBER: US/09/606,421B
;; NUMBER OF SEQ ID NOS: 358
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO: 153
;; LENGTH: 2007
;; TYPE: DNA
;; ORGANISM: Homo sapien
US-09-606-421B-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGAGTATGATCAGCTGTTACCGGTTGTTACCATGAATGCAATGCACCTGC 178
DB 1437 GCTTGTTCAATGAACACCATGTTCCCGAGAAAATCTCTACAAAGATGTTCAATTTT 1378
QY 179 AAGCGATGCGCCGTAATCTGAAGACCATTAACCCAGAGCGGTTATGTTAGCGCCG 238
DB 1377 AGTTAAAGTACTTAAAGTTCACTGGAAATTTTCAATGCAATGCAAGAAACTGGAGCT 1318
QY 239 ATGTAGCCAGGGTGGGATTAATAGTACTATTAAGTCACTGAGATGTTGAGAAATA 298
DB 1317 GCCTAGGGGAATTTTCAGATTAGGCTTGAATTACTCAACCCAGAGAAATCCAGATA 1258
QY 299 ATGCCACATCGACCAAGTGAAGCTTAAAACTCCGA 335
DB 1257 ATCTGCCCTCCCATTTAGTGTAGAGAGAGAGCTCCAA 1221

RESULT 15
US-09-221-107-153/c
;; Sequence 153, Application US/09221107
;; Patent No. 6660838
;; GENERAL INFORMATION:
;; APPLICANT: Wang, Tonglong
;; APPLICANT: Fan, Liqun
;; APPLICANT: Kalos, Michael D.
;; APPLICANT: Bangur, Chaitanya S.
;; APPLICANT: Hosken, Nancy
;; APPLICANT: Fanger, Gary R.
;; APPLICANT: Li, Samuel X.
;; APPLICANT: Wang, Aijun
;; APPLICANT: Skeiky, Yasir A.W.
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF LUNG CANCER
;; FILE REFERENCE: 210121.455C2
;; CURRENT APPLICATION NUMBER: US/09/221,107
;; NUMBER OF SEQ ID NOS: 161
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO: 153
;; LENGTH: 2007
;; TYPE: DNA
;; ORGANISM: Human
US-09-221-107-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGAGTATGATCAGCTGTTACCGGTTGTTACCATGAATGCAATGCACCTGC 178

```

Db      1437 GCTTGTTCATTGACACCATGTTCCCGAAGAAATTCCTACAAAGATGTTCAATTTT 1378
Qy      179  AAAGCATGCCCGTAAATCTGAAGACCATTAACCGAGCGGTTATGTAACGGCGCG 238
Db      1377 AGTTAAAGTACTTAAAGTTCACTGCGAATTTGCAATGACAGTGAAGAAAACGAGCT 1318
Qy      239  ATGTAGGCGAGGCGGTGGATTAATAGTACTATTGAATGACTGAGAAATGTTTCAAGAAATA 298
Db      1317 GCCTTAGGGGAATTTCAAGCAATTAGGCTTGAAATTACTCACCGAGAGAAATCCAGTATA 1258
Qy      299  ATGCCACCATCGACCGAGTGAACGCTAAATACTCCGA 335
Db      1257 ATCTGCCCTCCCATTAAGTGTAGAGAGAAAGCTCCA 1221

```

Search completed: March 16, 2004, 04:37:43
 Job time : 60.6647 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgc.....ccacgcctaccagatratraa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.4	8.0	2610	US-10-282-122A-30140	Sequence 30140, A
C 2	36.6	7.9	2667	US-10-282-122A-31185	Sequence 31185, A
C 3	35.8	7.9	954	US-10-282-122A-13780	Sequence 13780, A
C 4	34.2	7.5	1355	US-10-045-674-593	Sequence 593, App
C 5	34.2	7.5	2547	US-10-282-122A-36154	Sequence 36154, A
C 6	34	7.5	1878	US-10-369-493-43934	Sequence 43934, A
C 7	34	7.5	2886	US-09-801-368-131	Sequence 131, App
C 8	33.6	7.4	2301	US-09-815-242-4844	Sequence 4844, A
C 9	33.6	7.4	2301	US-09-815-242-9042	Sequence 9042, App
C 10	33.4	7.3	3140	US-10-425-114-29450	Sequence 29450, A
C 11	33.4	7.3	3485	US-10-434-599-64614	Sequence 64614, A
C 12	33.2	7.3	493	US-10-282-122A-26087	Sequence 26087, A
C 13	33.2	7.3	992	US-10-425-114-31747	Sequence 31747, A
C 14	32.8	7.2	568	US-10-333-631-1	Sequence 1, Appl1
C 15	32.8	7.2	718	US-10-027-632-26874	Sequence 26874, A

C 16	32.8	7.2	1059	US-09-738-626-482	Sequence 482, App
C 17	32.8	7.2	1113	US-09-738-626-481	Sequence 481, App
C 18	32.8	7.2	3309400	US-09-738-626-1	Sequence 1, Appl1
C 19	32.6	7.1	960	US-10-282-122A-11889	Sequence 11889, A
C 20	32.4	7.1	789	US-10-156-761-2022	Sequence 2022, App
C 21	32.4	7.1	9025608	US-10-156-761-1	Sequence 1, Appl1
C 22	32.2	7.1	972	US-10-333-631-4	Sequence 4, Appl1
C 23	32.2	7.1	1015	US-10-198-486-4377	Sequence 4377, App
C 24	32.2	7.1	1083	US-10-369-493-33133	Sequence 33133, A
C 25	32.2	7.1	4830	US-10-282-122A-27869	Sequence 27869, A
C 26	32	7.0	669	US-10-282-122A-31444	Sequence 31444, A
C 27	32	7.0	1001	US-10-425-114-14270	Sequence 14270, A
C 28	32	7.0	1152	US-10-156-761-303	Sequence 303, App
C 29	32	7.0	9025608	US-10-156-761-1	Sequence 1, Appl1
C 30	31.8	7.0	678	US-10-156-761-1219	Sequence 1219, App
C 31	31.8	7.0	930	US-09-815-242-6028	Sequence 6028, App
C 32	31.8	7.0	930	US-10-282-122A-20335	Sequence 20335, A
C 33	31.8	7.0	930	US-10-369-493-24480	Sequence 24480, A
C 34	31.8	7.0	978	US-10-369-493-24265	Sequence 24265, A
C 35	31.8	7.0	2766	US-10-282-122A-42042	Sequence 42042, A
C 36	31.6	6.9	951	US-10-369-493-43201	Sequence 43201, A
C 37	31.6	6.9	1338	US-10-369-493-37494	Sequence 37494, A
C 38	31.6	6.9	1642	US-09-887-576-469	Sequence 469, App
C 39	31.4	6.9	734	US-10-425-114-566	Sequence 566, App
C 40	31.4	6.9	1320	US-09-738-630-70	Sequence 70, Appl1
C 41	31.4	6.9	1467	US-10-282-122A-25701	Sequence 25701, A
C 42	31.4	6.9	1833	US-09-989-920-87	Sequence 87, Appl1
C 43	31.4	6.9	2007	US-09-735-705-153	Sequence 153, App
C 44	31.4	6.9	2007	US-09-850-716A-153	Sequence 153, App
C 45	31.4	6.9	2007	US-09-897-778-153	Sequence 153, App

ALIGNMENTS

RESULT 1
US-10-282-122A-30140/c
Sequence 30140, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyklich, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EITRA.034A
CURRENT APPLICATION NUMBER: US/10/282, 122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191, 078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206, 848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207, 727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230, 335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230, 347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242, 578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253, 625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257, 931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267, 636

PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 30140
LENGTH: 2610
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-30140

Query Match
Best Local Similarity 8.0%; Score 36.4; DB 12; Length 2610;
Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 17 TGGCAGCATTCGAGCAATCGTAGTTCTGCGAGTCTCTGCGCTGCGCTTCCCAAT 76
DB 410 TTGGCGCGCTCGCGCGCGCTCGGAAAGCGCTGCAATCAGCAGGGTGTCTTCACTGT 351
QY 77 GGGCGCGCGCGGTATCATTAACGGCGCGCGCAATAGTTCCGGCCCGCACTATGATCAGC 136
DB 350 TCGCGCGCGGTGTGGCCAGTTTGGCGCTCTCCAGCACTTGCACAGGGTGGCGATGTTG 291
QY 137 TGGTAAACCGGTGTGTTACCAATGCAATGCACTGCAAAAGCGATGCCGTAAT 196
DB 290 TGGCGCGCTGATGATGTCGCCAGCAGCTCGCGCGCACTGTTGTGCAATAGGGGGGAG 231
QY 197 CTGAACACCAATTAACCAAGCGGTATGATTAACGGCGCGCGCAATAGGC 246
DB 230 GTGGCTTGGCCCTTGACGATGCGGAAAGATCCGGCTTGCATGAGGC 181

RESULT 2

US-10-282-122A-33185/c
Sequence 33185, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 33185
LENGTH: 2667
TYPE: DNA
ORGANISM: Pseudomonas syringae
US-10-282-122A-33185

Query Match
Best Local Similarity 7.9%; Score 36; DB 12; Length 2667;
Matches 159; Conservative 0; Mismatches 205; Indels 0; Gaps 0;

QY 25 TTGCGACCAATCGTAGTTCTGCGAGTCTCTGCGCTGCGCTTCCCAATGCGCGC 84
DB 471 TTGGCGCGCTCGCGCGCGCTCGGAAAGCGCTGCAATCAGCAGGGTGTCTTCACTGT 412
QY 85 GCGGTATCATTAACGCGCGCGCAATAGTTCCGGCCCGCACTATGATCAGTGTACC 144
DB 411 GACCGCGCGCGGTGTGGCCAGTTTGGCGCTCTCCAGCACTTGCACAGTGTGATGCGCGC 352
QY 145 CGTGTGTTAACCATGAATGCAATGCACTGCAAAAGCGATGCCGTAATCTGAACG 204
DB 351 CTGCATGTGTCCAGAGCTCGTGTGCGCTTTTGTGCAATCAGGGGAGAGTGTGCTT 292
QY 205 ACCATTACCAAGCGGTATGATTAACCGCGCGCATGATGAGCGCGGTGCGATATAGT 264
DB 291 GCCCTTGGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 232
QY 265 ACTATTGAATGATCTGAGATGTTTCAAGAAATATCCATCGACAGCGTGAACGCT 324
DB 231 TGAATGCGATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 172
QY 325 AAAAATCGCATTTATCTGCGCAATTAACGGCGGTATTAACGGCGGTGATATCAG 384
DB 171 CAGCAGCTGACGAGGCGTGGTGTGTTGCGGTTTAAAGCGGTGAGGAGGAGGAGGAGGAG 112
QY 385 ACCG 388
DB 111 GCGC 108

RESULT 3

US-10-282-122A-13780/c
Sequence 13780, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
PRIOR FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO: 13780
LENGTH: 954
TYPE: DNA
ORGANISM: Burkholderia mallei
US-10-282-122A-13780

Query Match 7.9%; Score 35.8; DB 12; Length 954;
Best Local Similarity 54.1%; Pred. No. 0.2;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 55 CTGGCTGCTGCTTCCACATGCGCGCGGCGGATTAATCAATGCGCGCGCAATAGT 114
Db 300 CTGGCGCCCATAGCGGAAGTAGGCGATGCGCGGTATCCGCGCGCGATGCGCGCTT 241
Qy 115 TCCGCGCGGACATGATGATGAGTGTGTTACCGGTGTTACCAATGAATGCGCATGCA 174
Db 240 GAGCGCATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 181
Qy 175 CTGCAAGCGATGCC 189
Db 180 TTGCAGACGAGAAC 166

RESULT 4

US-10-045-674-593
Sequence 593, Application US/10045674
Publication No. US2003023333A1
GENERAL INFORMATION:
APPLICANT: LADNER, ROBERT C.
APPLICANT: COHEN, EDWARD H.
APPLICANT: NASTRI, HORACIO G.
APPLICANT: ROONEY, KRISTIN L.
APPLICANT: HOEFT, RENE
TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
TITLE OF INVENTION: LIBRARIES
FILE REFERENCE: DVAX/002 CIP2
CURRENT FILING DATE: US/10/045, 674
PRIOR APPLICATION NUMBER: 60/198,069
PRIOR FILING DATE: 2000-04-17
PRIOR APPLICATION NUMBER: 09/837,306
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 635
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 593
LENGTH: 1355
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: M13-III
FEATURE:
NAME/KEY: CDS
LOCATION: (1) : (1305)
US-10-045-674-593

Query Match 7.5%; Score 34.2; DB 15; Length 1355;
Best Local Similarity 26.9%; Pred. No. 0.86; Indels 169; Gaps 0;
Matches 90; Conservative 76; Mismatches 169; Indels 0; Gaps 0;

Qy 20 CAGATTGCGAGCATGTAATTTCTGCGAGTGTCTGCTGCGCTTCCACATGCG 79
Db 755 CMCCKGTAAVGCYAGCGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGG 814
Qy 80 GCGCGCGCGGTAATCATTAACGCGCGCGCAATAGTTCGCGCGCGCATATGATGACTGCG 139
Db 815 GYMSYGARGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGG 874
Qy 140 TTACCGGTGTTTACCATGAAATGGACATGACATGACATGACATGACATGACATGACATG 199
Db 875 TGGCMAATGCTAAATGAGGSCATGACATGACATGACATGACATGACATGACATGACATG 934
Qy 200 AAGCACCATTACCCAGAGCGGTTATGTTAAGCGCGCGCATGATGACCGAGGATGCGGATA 259
Db 935 AYGCVAAAGGYAATYGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 994
Qy 260 ATAGTACTATTAATGAACTGACATGATGATGATGATGATGATGATGATGATGATGATG 319
Db 995 TYGGYAGYGTGYSYGGYCTKCTTAAGGYAAGGAGYAGGAGYAGGAGYAGGAGYAGGAGY 1054
Qy 320 ACGTAAACTCGGATATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 354
Db 1055 ATTCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1089

RESULT 5

US-10-282-122A-36154
Sequence 36154, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT FILING DATE: US/10/282, 122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1

	Query Match	7.3%;	Score 33.4;	DB 12;	Length 3140;
	Best Local Similarity	45.8%;	Pred. No.2.6;	Mismatches 0;	Gaps 0;
	Matches 115;	Conservative	0;	Indels 136;	Gaps 0;
Qy	91 AATCATTAACGGCGCGGCAATAGTTCCGCGCCGACATATGATTCAGCTGTTACCCGCTTT				
Db	1801 AATTAAGGCAAGGCCCTTAATAGGCTTAATCAGTAAATGATTAAGGCATAGCTGGTAA				
Qy	151 GTTACCACATGAATGCGACATGCACCTGCAAGCGAGTSCCGTAAATCTGAACGACATT				
Db	1861 ACTGCATCTGGAGAGCTAGAGAAATGGAACAAGATGCTCAGAAATGTTGAACAAGTGT				
Qy	211 ACCCAGACGGTTATGTGTAACGCGCCGAGTGTAGCCAGGTCGGGATTAATAGTACTATT				
Db	1921 TCAATATGTATGTTCTGCAAAAGAAAGAAATGTTTGTATATATATGATTAATAGAGAGCTT				

FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 31747
LENGTH: 992
TYPE: DNA
ORGANISM: Zee mays
FEATURE:
OTHER INFORMATION: Clone ID: UC-2MFLB/3212F11_F11
US-10-425-114-31747

Query Match 7.3%; Score 33.2; DB 12; Length 992;
Best Local Similarity 52.1%; Pred. No. 1.6;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 19 GCAGCATTCGACGACATGCTAGTCTCTGACAGTCTGCTGCGCTGCTTCCACATG 78
Db 497 GCTGAGCGGATGACGAGCTTGTGCGGGGTGAGGTCTCGGGGCTTGTGTCTCGGC 438
Qy 79 GCGCGCGCGGTATCATATACGCGCGGCAATAGTTCGCGCGGCACTATGATGACTG 138
Db 437 GCGCGCGCGGTATCATATACGCGCGGCAATAGTTCGCGCGGCACTATGATGACTG 378
Qy 139 GTTACCCGCTGTTGTTACCATG 160
Db 377 GCGAGCGGTGCGCTTGGCGGTG 356

RESULT 14

US-10-333-631-1
Sequence 1, Application US/10333631
Publication No. US20030190651A1
GENERAL INFORMATION:
APPLICANT: Bayer AG
TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
FILE REFERENCE: L10121 Foreign Countries
CURRENT FILING DATE: 2003-01-23
PRIOR APPLICATION NUMBER: US/10/333,631
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/221,917
PRIOR FILING DATE: 2001-04-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (90)..(90)
OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

Query Match 7.2%; Score 32.8; DB 14; Length 568;
Best Local Similarity 51.0%; Pred. No. 1.6;
Matches 76; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

Qy 139 GTTACCCGCTGTTGTTACCATGAAATGACATGCACTGCAAGGATGCCGTAAATCT 198
Db 87 GTTCGAGCTGCTGACAGCTGCTGAAACAGGCTTGGCTGCACTCAAGCCCTGTCGCT 146
Qy 199 GAAAGACATTATCCACAGAGCGGTATGTAACGCGCGCATGTAGCCAGAGTCCGAT 258
Db 147 GAAATGCTGCTCAACGAGCACTGAGCTGACATGACAGAGAGCTGATGCGGGTGCCTA 206
Qy 259 AATAGTACTATTGAAGTACTGAGATG 287
Db 207 CCGATCTGCGGATAGCCACATGCTGG 235

RESULT 15

US-10-027-632-26874
Sequence 26874, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 26874
LENGTH: 718
TYPE: DNA
ORGANISM: Human
US-10-027-632-26874

Query Match 7.2%; Score 32.8; DB 15; Length 718;
Best Local Similarity 49.4%; Pred. No. 1.9;
Matches 85; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 18 GCGACATTCGACGACATGCTAGTCTTGTGCGAGTCTGCTGCGCTGCTTCCACATG 77
Db 514 GAAAGATCCACACACACAAAGTGCCTGCGCTGCTGCTGCTGCTGCTGCTGCTG 573
Qy 78 GCGCGCGCGGTATCATATACGCGCGGCAATAGTTCGCGCGGCACTATGATGACT 137
Db 574 GCTGCGAGCTGCCACAGCTGTCGGGCTGAGCTTCCGTCCTGCGATCTACAGC 633
Qy 138 GTTACCCGCTGTTGTTACCATGAAATGACATGCACTGCAAGGATGCC 189
Db 634 TGTGTCCTGTAAGCAGACATGACTGGGCAACGAGCGGGGTTCTATGCC 685

Search completed: March 17, 2004, 08:16:20
Job time : 413.736 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgc.....caccgctaacagctatcaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_estc3:*
12: gb_estc4:*
13: gb_estc5:*
14: gb_estc6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
C 1	232.6	51.0	680 12	Bu618688 Bu618688
C 2	37.4	8.2	484 9	AVA30994 AVA30994
C 3	37	8.1	590 14	CF848025 CF848025
C 4	37	8.1	610 14	CF842541 CF842541

Result No.	Score	Query Length	ID	Description
C 5	36.6	8.0	891 29	CNS02189
C 6	36.4	8.0	787 28	BZ568375
C 7	35.2	7.7	417 29	FR0020317
C 8	35.2	7.7	811 28	BZ469000
C 9	35.2	7.7	813 28	BZ204853
C 10	35	7.7	490 14	CAS88919
C 11	34.8	7.6	480 9	AU085223
C 12	34.6	7.6	406 9	AU241511
C 13	34.6	7.6	499 12	BJ006097
C 14	34.6	7.6	561 12	BJ001420
C 15	34.6	7.6	640 12	BJ000263
C 16	34.6	7.6	656 12	BJ004146
C 17	34.6	7.6	662 12	BJ010653
C 18	34.6	7.6	662 12	BJ011743
C 19	34.6	7.6	671 12	BJ012818
C 20	34.6	7.6	682 14	CF402142
C 21	34.6	7.6	691 12	BJ000180
C 22	34.6	7.6	702 12	BG441624
C 23	34.6	7.6	937 12	BG444255
C 24	34.4	7.5	365 9	AA746477
C 25	34.4	7.5	500 14	CA711477
C 26	34.4	7.5	827 28	BH438448
C 27	34.2	7.5	296 9	AV058630
C 28	34.2	7.5	523 12	BJ334624
C 29	34	7.5	588 14	CA731849
C 30	33.8	7.4	522 14	CA707989
C 31	33.8	7.4	625 13	BU882267
C 32	33.8	7.4	1098 29	CNS02GQP
C 33	33.6	7.4	584 28	AQ398232
C 34	33.6	7.4	594 12	BJ335653
C 35	33.6	7.4	653 12	BJ336964
C 36	33.6	7.4	712 14	CF437982
C 37	33.6	7.4	866 29	CNS02C24
C 38	33.6	7.4	896 28	CC33807
C 39	33.6	7.4	910 14	CA266034
C 40	33.4	7.3	525 9	AU242732
C 41	33.4	7.3	596 9	AU177600
C 42	33.4	7.3	649 12	BI261742
C 43	33.4	7.3	679 29	CC961526
C 44	33.4	7.3	695 28	BH63951
C 45	33.4	7.3	700 28	BH651371

ALIGNMENTS

RESULT 1
LOCUS Bu618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION Bu618688 NIBB Mochii normalized Xenopus early gastrula library
ACCESSION Bu618688
VERSION Bu618688.1 GI:37256713
KEYWORDS
SOURCE
ORGANISM Xenopus laevis (African clawed frog)
EST

REFERENCE
AUTHORS Kikayota, M., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
1 (bases 1 to 680)
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp
The information of this clone is available through the following URL.

http://xenopus.nibb.ac.jp.

FEATURES Location/Qualifiers

source 1..680
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="X1186D22"
 /tissue_type="whole embryo"
 /dev_stage="stage 10.5"
 /clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 51.0%; Score 232.6; DB 12; Length 680;
 Best Local Similarity 69.3%; Pred. No. 4.7e-54;
 Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

OY 1 ATGAACCTTTAAAGTGGAGCATTCGACGACATCGTAGTTCTGGGAGTGGCTGGCT 60
 Db 593 ATGAACCTTTAAAGTGGAGCATTCGACGACATCGTAGTTCTGGGAGTGGCTGGCT 60
 OY 61 GGGCTGCTCCACATAGGGGCGGCGGCGGTATCTAAGCGGCGCAATAGTCCGCG 120
 Db 533 GGTGTGTTCTCAGTAGCGCGCGCGGTATCCACGGTGTGGCGGTAAATAGCGGC 474
 OY 121 CCGGACTATGATCAGTGTGTACCGGTGTGTGTACCCATGAAATGGACATGACCTGCA 180
 Db 473 CCAATTTGACGTGAACATTTTACCAATGACGAGCGGTAACTNTGCACTTCTCTGCA 414
 OY 181 AGCGATGCCCTGTAATCTGAAACGACATTCACGAGCGGTATGTAAAGCGCGCAT 240
 Db 413 ACTGATGCCCGTAACTCTGACTTGAATCCAGAGTGGCGCGGTATGTGTCAAGAT 354
 OY 241 GTAGGCCAGGAGTGGCAATATGATCTAATGAACTGACATGATGTTTCAGAAATAT 300
 Db 353 GTGTGTGAGGCTCAGATGACGCTCAATGATGATCTGACCAAGTGGCTTCGTTAAGC 294
 OY 301 GCCACATCGACAGTGGAAAGCTCAAAATCTCGATATTAATCTGGCCCAATAGCGCGT 360
 Db 293 GCTACTCTTATCATGTGAGACGGCAAAATCTGAAATGACGTTAAACAGTTCGGTGT 234
 OY 361 AATACGCCCGCTGGTATATCAGACCGCATCTGATTCAGAGTATGATGTCGTCAGT 420
 Db 233 GGCACGGTGTCTCAATGACGATGATGATGATGATGATGATGATGATGATGATGAT 174
 OY 421 GGTTCGCAACAGCGCAAGCGCTAATGATGATGATGATGATGATGATGATGATGAT 456
 Db 173 GGCCTTGTAAACACCGACCGCTCATGATGATGATGATGATGATGATGATGATGAT 138

RESULT 2
 AV430994/c

LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
 DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
 ACCESSION AV430994
 VERSION AV430994.1 GI:8586219
 KEYWORDS EST.
 SOURCE Porphyra yezoensis
 ORGANISM Porphyra yezoensis
 Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
 Porphyra

REFERENCE 1 (bases 1 to 484)
 AUTHORS Nakaio, T., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
 TITLE Generation of 10,154 expressed sequence tags from a leafy
 JOURNAL gametophyte of a marine red alga, Porphyra yezoensis
 MEDLINE DNA Res. 7, 223-227 (2000)
 PUBMED 20363100
 COMMENT 10907854
 Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 Location/Qualifiers

FEATURES source 1..484
 /organism="Porphyra yezoensis"
 /mol_type="mRNA"
 /strain="TU-1"
 /db_xref="taxon:2788"
 /clone="PL027e10.r"
 /clone_lib="Porphyra yezoensis TU-1"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
 Best Local Similarity 57.1%; Pred. No. 15;
 Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

OY 321 CGTAAACTCCGATATTACTGTGCGCAATAGCGGATATTAACCGCGCTGTTAA 380
 Db 441 CCAAAACACAAAGTGTCCGCGCTCAAGACGCGCTCAATACCGCTGCGCAG 382
 OY 381 TCAGACCGATCTGATTCAGCGTATGTGTGTCAGTGTGTTGGCAACAGCGCA 439
 Db 381 GCCGAAGCTGTGTATCAAGTACTGTGTCAGGTTGCGGAAAGGCGTCCCGCA 323

RESULT 3

CF848025 590 bp mRNA linear EST 30-OCT-2003
 LOCUS CF848025
 DEFINITION pSMA003x114f USDA-IPAFS: Expression of Phytophthora sojae genes
 during infection and propagation_sma Phytophthora sojae cDNA clone
 ACCESSION CF848025
 VERSION CF848025.1 GI:38063679
 KEYWORDS EST.
 SOURCE Phytophthora sojae
 ORGANISM Phytophthora sojae
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 Phytophthora

REFERENCE 1 (bases 1 to 590)
 AUTHORS Tyler, B.
 TITLE Tyler, B. Not Published
 JOURNAL Unpublished (2003)
 COMMENT Contact: Tyler B
 Tyler lab

VI

1880 Pratt Dr., Blacksburg, VA 24061, USA
 Tel: 540-231-7318
 Email: bmyler@vt.edu
 PCR PRIMER
 FORWARD: BK reverse primer
 BACKWARD: BK reverse primer
 Plate: 003 row: 1 column: 14
 Seq primer: BK reverse primer
 High quality sequence stop: 590.
 Location/Qualifiers

FEATURES

source 1..590
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="SMA003114"
 /tissue_type="mycelium"
 /cell_line="P6497"
 /dev_stage="mycelium"
 /lab_host="synthetic medium"
 /clone_lib="USDA-IPAFS: Expression of Phytophthora sojae
 genes during infection and propagation_sma"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 590;
 Best Local Similarity 51.5%; Pred. No. 21;

Matches	85; Conservative	0; Mismatches	80; Indels	0; Gaps
Qy	26	TCGACGCAATGTAAGTATCTTGTGCGAGTCTCTGCGCGGTGCTTCCACAAATGGGCGGCG	85	
Db	224	TTGTAGCGCGTAGTCCCTTGATGTCCGTGAGCTTGGGGAAGTTCGCCGATGAAATGAGCGGCG	283	
Qy	86	GCGGTAATCATPAAGCGCGCGGCATATGTTCCGGCCCGGACTATGATACGCTGTTACCC	145	
Db	284	CCGTTGATGAGCAGACAGCGCGCCCAAAAGCCGTCGCGCGGATATGAGCCAGCGCTCG	343	
Qy	146	GTGTTGTTACCCATGAATGGCAGCATGCACTGCAACAAAGATGCC	190	
Db	344	ATCTTGAAAGCCAGCAGCGCGCGGATGAGCCCGCCACAGATGCGC	388	
RESULT 4				
LOCUS	CF842541			
DEFINITION	CF842541	610 bp	mRNA	EST 30-OCT-2003
ACCESSION	CF842541			
VERSION	CF842541			
KEYWORDS	EST.			
SOURCE	Phytophthora sojae			
ORGANISM	Phytophthora sojae			
REFERENCE	Eukaryota; Eimeriopsis; Oomycetes; Pythiales; Pythiaceae;			
AUTHORS	Phytophthora sojae			
TITLE	Phytophthora.			
JOURNAL	1 (bases 1 to 610)			
COMMENT	Tyler B. Not Published			
	Unpublished (2003)			
	Contact: Tyler B			
	Tyler lab			
	VBI			
	1880 Pratt Dr., Blacksburg, VA 24061, USA			
	Tel: 540-231-7318			
	Email: bmttyler@vt.edu			
	PCR Primers			
	FORWARD: BK reverse primer			
	BACKWARD: BK reverse primer			
	Plate: 020 row: G column: 21			
	Seq primer: BK reverse primer			
	High quality sequence stop: 610.			
FEATURES				
source				
	1..610			
	/organism="Phytophthora sojae"			
	/mol_type="mRNA"			
	/db_xref="taxon:67593"			
	/clone="SHB020G21"			
	/tissue_type="mycelium"			
	/cell_line="P6497"			
	/dev_stage="48 hr. post infection stage"			
	/lab_host="Soybean plant"			
	/clone_id="USDA-IPAFs:Expression of Phytophthora sojae			
	genes during infection and propagation_SHB"			
	/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"			
ORIGIN				
Query Match	8.1%	Score 37;	DB 14;	Length 610;
Best Local Similarity	51.5%;	Pred. No. 22;		
Matches	85; Conservative	0; Mismatches	80; Indels	0; Gaps
Qy	26	TCGACGCAATGTAAGTATCTTGTGCGAGTCTCTGCGCGGTGCTTCCACAAATGGGCGGCG	85	
Db	212	TTGTAGCGCGTAGTCCCTTGATGTCCGTGAGCTTGGGGAAGTTCGCCGATGAAATGAGCGGCG	271	
Qy	86	GCGGTAATCATPAAGCGCGCGGCATATGTTCCGGCCCGGACTATGATACGCTGTTACCC	145	
Db	272	CCGTTGATGAGCAGACAGCGCGCCCAAAAGCCGTCGCGCGGATATGAGCCAGCGGCTCG	331	
Qy	146	GTGTTGTTACCCATGAATGGCAGCATGCACTGCAAAAGATGCC	190	

RESULT 5	CNS02189/c	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	PUBMED	REFERENCE	AUTHORS	COMMENT	FEATURES	ORIGIN
CNS02189	891 bp	DNA	linear	GSS 01-SEP-2000															
Tetradon nigroviridis	genome survey sequence	PUC-Or1	end of clone																
141g11 of library G	from Tetradon nigroviridis	genomic survey																	
sequence.																			
GI:7836745																			
GSS: genome survey sequence.																			
Tetradon nigroviridis																			
Tetradon nigroviridis																			
Euhariyola, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;																			
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;																			
Tetraodontiformes;																			
Tetraodonidae; Tetraodon.																			
Ruest Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,																			
Bernot, A., Fitzames, C., Wincker, P., Brothier, P., Quetier, F.,																			
Saurin, W. and Weissenbach, J.																			
Characterization and repeat analysis of the compact genome of the																			
freshwater pufferfish Tetradon nigroviridis																			
Genome Res. 10 (7), 939-949 (2000)																			
10899143																			
3 (bases 1 to 891)																			
Genoscope.																			
Direct Submission																			
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :																			
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr																			
- Web : www.genoscope.cns.fr)																			
This sequence is a single read and was generated as part of a large																			
scale clone-end sequencing project of the Tetradon nigroviridis																			
genome. For more information, please take a look at																			
http://www.genoscope.cns.fr/tetradon.																			
Location/Qualifiers																			
1..891																			
/organism="Tetradon nigroviridis"																			
/mol_type="genomic DNA"																			

RESULT 6
 BZ568375
 LOCUS
 DEFINITION BZ568375 787 bp DNA linear GSS 17-DEC-2002
 accession BZ568375_1 pac62-164_7503, genomic survey sequence.
 VERSION BZ568375.1 GI:27201285
 KEYWORDS GSS.
 SOURCE Pseudomonas aeruginosa
 ORGANISM Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.
 REFERENCE 1 (bases 1 to 787)
 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
 Burns,J.L., Kaul,R. and Olsen,M.V.
 Whole-Genome-Sequence variation among multiple isolates of
 Pseudomonas aeruginosa library
 J. Bacteriol. (2002) in press
 Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
 Class: shotgun.
 FEATURES
 source
 Location/Qualifiers
 1..787
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="2-164"
 /db_xref="taxon:287"
 /clone="pac62-164_7503"
 /clone_lib="pac62-164"
 /note="clinical isolate 2-164 Whole genomic shotgun
 library."
 ORIGIN
 Query Match 8.0%; Score 36.4; DB 28; Length 787;
 Best Local Similarity 47.4%; Pred. No. 36;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
 QY 17 TGGAGAGATTCGACGACATGTTCTTGGCAGAGTCTGGCTGGCGCTCTTCACAAT 76
 Db 146 TTGGCCGGCTGGCGACGCTGTGGAAGGCTGCAACATCAGACAGGTGTCTTACGCTGT 205
 QY 77 GGGGCGCGCGCGGTATCATACGCGCGCGCAATAGTTCCGCGCGGACTATGATCAGC 136
 Db 206 TCGCGCGCGGTGTGGCCAGTTCGCGCTGTCCAGCACTGCACGAGGTGGCGATGTTG 265
 QY 137 TGGTACCGGTGTGTATCCCATGGAATGACATGCACTGCACAAAGGATGCCCGTAAT 196
 Db 266 TAGCCGCCCTGCATGTGTCCCGACGCTCGCGGACGTTGCTTGTGATAGGGGGAG 325
 QY 197 CTGAACGACATTAACCGAGCGGTATGTTACGCGCGCGATGTAGGC 246
 Db 326 GTGGCTTGGCCCTTGACGATGCGGAAAGAACCGGCTTGAAGTGGC 375

RESULT 7
 FR0020317
 LOCUS
 DEFINITION FR0020317 417 bp DNA linear GSS 10-DEC-1997
 accession AL013202 F.rubripes GSS sequence, clone 040D20BD10, genomic survey sequence.
 VERSION AL013202.1 GI:2679570
 KEYWORDS GSS; genome survey sequence.
 SOURCE Takifugu rubripes (Fugu rubripes)
 ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorphi; Tetraodontiformes;
 Tetraodontidae; Tetraodontidae; Takifugu.

REFERENCE 1 (bases 1 to 417)
 AUTHORS Elgar,G., Clark,M., Smith,S., Week,S., Warner,S., Umrania,Y.,
 Williams,G. and Brenner,S.
 TITLE Direct Substitution
 JOURNAL Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource
 Centre Hinxton, Cambridge, CB10 1SB. Email: biolhelp@hgm.mrc.ac.uk
 COMMENT Vector: phagescript II KS
 V type: phagemid
 PRIMER: KS
 DECKR: One pass dye-terminator sequencing of cosmid cloned genomic
 sequence.
 FEATURES
 source
 Location/Qualifiers
 1..417
 /organism="Takifugu rubripes"
 /mol_type="genomic DNA"
 /db_xref="taxon:31033"
 /clone="040D20BD10"
 /clone_lib="cosmid 040D20"

ORIGIN

Query Match 7.7%; Score 35.2; DB 29; Length 417;
 Best Local Similarity 50.3%; Pred. No. 59;
 Matches 76; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 170 ATGCACTGCAAGACGATGCCGTAAATCTGAAGACCATTTACCGAGCGTTATGTA 229
 Db 16 ATGCAATCCAAAGAGACACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75
 QY 230 ACCGCGCGCGATGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 75
 Db 76 ATTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 289
 QY 290 TGAAGAAATATGCAACCATGACCAAGTGA 320
 Db 136 AGAGAACGATTAATACACGACCTTAGGAA 166

RESULT 8
 BZ469000/C
 LOCUS
 DEFINITION BZ469000 811 bp DNA linear GSS 13-DEC-2002
 accession BZ469000 BOAO58TF_BO_1.6_2_KB_tot Brassica oleracea genomic clone BOAO58,
 genomic survey sequence.
 VERSION BZ469000.1 GI:26764546
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 811)
 Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 unpublished (2001)
 JOURNAL Other GSSs: BOAO58TF
 COMMENT Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
 FEATURES
 source
 Location/Qualifiers
 1..811
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOAO58"
 /clone_lib="BO_1.6_2_KB_tot"

ORIGIN /note="Vector: PHOS1, Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into PHOS1 using BstXI linkers"

Query Match 7.7%; Score 35.2; DB 28; Length 811;
Best Local Similarity 47.7%; Pred. No. 78;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 146 GTGTTTACCATTAAATGACATGACATGCAAGGATCCCGTAAATGTAAGCA 205
DB 640 GTATTAAATTAAGATGTTGTAAGCTTGATGTTGGAAGCATTTGTAACCTGTGTTTC 581
QY 206 CCATTACCCAGAGCGGTTATGTTAAGCGGCGCATGATAGCGGATGATTAATGTA 265
DB 580 CCATTTCATCTACTGTTTATGCAAAATGCTTGTGATCTAATTCATCTGATGTTGAGCTG 521
QY 266 CTATTGAACTGACATGCAAAATGTTTCAGAAATTAATGCAACCATGCAACGATGGAACGCTA 325
DB 520 CTCCTTGCTGCTGATGAGGATAGTTTAAATTAATCTTGTGACATTAATTAATGAAATGCA 461
QY 326 AAAATCCGATATTAATCTGTCGCGCAATACGGCGGTA 361
DB 460 AATACACCTGATTAACACTTATTAACCAACCAAGATA 425

RESULT 9 BZ204853 813 bp DNA linear GSS 11-OCT-2002
LOCUS BZ204853
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-390D21, genomic survey sequence.

ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

REFERENCE 1 (bases 1 to 813)
AUTHORS Zhao,S., Shetty,J., Shatsman,S., Tsengaye,G., Geer,K.,
Shvartsbeyn,A., Gebregiorgis,E., Overton,L., Russell,D., Chen,D.,
Riggs,F., de Jong,P. and Fraser,C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)

TITLE Other-GSS: CH230-390D21.T1
JOURNAL
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igf.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.html). For BAC library
availability, please contact Pieter de Jong (pje@jngmail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_ering_information.html). BAC end
pages: http://www.cligr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 390 row: D column: 21
Seq primer: T7
Class: BAC ends.

FEATURES
SOURCE Location/Qualifiers

1..813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_1lb="CHORI-230 Segment 2"
/note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SNHsd/MCW) BAC library produced by

ORIGIN Pieter de Jong"

Query Match 7.7%; Score 35.2; DB 28; Length 813;
Best Local Similarity 49.0%; Pred. No. 78;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 178 CAAGCGATGCGCCGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAGCGGCC 237
DB 548 CAAGACCTGCACATCAAAACAGACACACTCAAACTAATTAAGAAAGAACTAGGAGACA 607
QY 238 GATGAGGCGCAGGCGCGATTAATGATGATTAATGATGATGATGATGATGATGATGATGAT 297
DB 608 TCTGGAACACATGCGGACACTGGAATAAATTCCTGAACAAACCAATGCTTATGCTCT 667
QY 298 AATGCCACATGACCACTGGAACGCTTAATAAATCCGATTAATGATGATGATGATGATGATGAT 357
DB 668 AAGATCAAGATGACCAATGAGGATCTCATTAACGCAAGCTTCTGTAGGCAAGAGACAC 727
QY 358 GGTAAATACGCC 369
DB 728 TGTGTTAGGAC 739

RESULT 10 CA588919/c 490 bp mRNA linear EST 19-NOV-2002
LOCUS CA588919/c
DEFINITION hab51c12.y1 Fugu UT7 adult skin Takifugu rubripes cDNA clone
IMAGE:6352126 5' similar to SW:CID4_MOUSE O35054 CLAUDIN-4 ; mRNA
sequence.

ACCESSION CA588919
VERSION CA588919.1 GI:25133497
KEYWORDS EST.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes

REFERENCE 1 (bases 1 to 490)
AUTHORS Clark,M., Johnson,S.L., Leirach,H., Lee,R., Li,F., Marra,M.,
Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schuck,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
WashU Zebrafish EST Project 1998
Unpublished (1998)

TITLE Washington University School of Medicine
JOURNAL
COMMENT Contact: Stephen L. Johnson
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu

Library materials provided by G. Elgar (UK MRC HGMP-RC) Library
constructed by Drs. K. Kawakami, M. Saeki, S. Sugano, K. Kikuchi
and S. Matsuda (University of Tokyo, Institute of Medical Science
and Laboratory of Aquatic Molecular Biology and Biotechnology) DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution: Fugu clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL, send email to:
info@image.jnl.gov
Seq primer: T3 ET from Amersham.

FEATURES
SOURCE Location/Qualifiers

1..490
/organism="Takifugu rubripes"
/mol_type="mRNA"
/db_xref="taxon:31033"
/clone="IMAGE:6352126"
/sex="Female"
/tissue_type="skin"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1lb="Fugu UT7 adult skin"

ORIGIN

Index	Gaps	0;
1	0;	0;
2	0;	0;
3	0;	0;
4	0;	0;
5	0;	0;
6	0;	0;
7	0;	0;
8	0;	0;
9	0;	0;
10	0;	0;
11	0;	0;
12	0;	0;
13	0;	0;
14	0;	0;
15	0;	0;
16	0;	0;
17	0;	0;
18	0;	0;
19	0;	0;
20	0;	0;
21	0;	0;
22	0;	0;
23	0;	0;
24	0;	0;
25	0;	0;
26	0;	0;
27	0;	0;
28	0;	0;
29	0;	0;
30	0;	0;
31	0;	0;
32	0;	0;
33	0;	0;
34	0;	0;
35	0;	0;
36	0;	0;
37	0;	0;
38	0;	0;
39	0;	0;
40	0;	0;
41	0;	0;
42	0;	0;
43	0;	0;
44	0;	0;
45	0;	0;
46	0;	0;
47	0;	0;
48	0;	0;
49	0;	0;
50	0;	0;
51	0;	0;
52	0;	0;
53	0;	0;
54	0;	0;
55	0;	0;
56	0;	0;
57	0;	0;
58	0;	0;
59	0;	0;
60	0;	0;
61	0;	0;
62	0;	0;
63	0;	0;
64	0;	0;
65	0;	0;
66	0;	0;
67	0;	0;
68	0;	0;
69	0;	0;
70	0;	0;
71	0;	0;
72	0;	0;
73	0;	0;
74	0;	0;
75	0;	0;
76	0;	0;
77	0;	0;
78	0;	0;
79	0;	0;
80	0;	0;
81	0;	0;
82	0;	0;
83	0;	0;
84	0;	0;
85	0;	0;
86	0;	0;
87	0;	0;
88	0;	0;
89	0;	0;
90	0;	0;
91	0;	0;
92	0;	0;
93	0;	0;
94	0;	0;
95	0;	0;
96	0;	0;
97	0;	0;
98	0;	0;
99	0;	0;

```

VERSION
AU085223.1
GT:9963206

```

Eukaryota; Viridiplantae; Streptophyta: Tracheophyta.

REFERENCE 1 (bases 1 to 480)

TITLE Expression analysis of ESTs derived from the inner bark of

PUBMED 11052197

FEATURES

Location/Qualifiers

[illegible]

latipes cDNA clone OLC53.01d similar to pif/T16763/hypothetical

VERSION AU241511.1
 CI:18154090

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS
Naruse, K., Mitani, H. and Tanaka, M.

Department of Biological Sciences

Source	Location/Qualifiers
1	

```
/organism="Oryzias latipes"
```

ORIGIN

[illegible][illegible]

Y 171 TGCAC TGCA 179

DEFINITION BJ006097 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA08BD11 5',
RNA sequence.
ACCESSION BJ006097
VERSION BJ006097.1 GI:17359814
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 499)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..499
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSA08BD11"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

ORIGIN
Query Match 7.6%; Score 34.6; DB 12; Length 499;
Best Local Similarity 54.3%; Pred. No. 94;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 51 TGCCTGCGCGCGCTGCTTCACAAATGGGCGCGCGGTATCATTAACGGCGCGGCA 110
DB 195 TGTCTCTGCGCGCGGTACCGGAAGACGGGAGATGACCGCGGTGACCGGACGACGCGGCA 136
OY 111 TAGTTCGGCGCGGACTATGATCAGCTGTACCGGTGTATCCATGGAATGGCACA 170
DB 135 TCTTGACAGCGGTGATCATGATCAGACACCTCATGAGTGTGATCTCACCAGCAGC 76
OY 171 TGCACCTGCA 179
DB 75 CGCTCTGCA 67

RESULT 14
BJ001420/c 561 bp mRNA linear EST 05-DEC-2001
LOCUS BJ001420 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA007C02 5',
DEFINITION mRNA sequence.
ACCESSION BJ001420
VERSION BJ001420.1 GI:17361687
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 561)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..561
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSA007C02"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

ORIGIN
Query Match 7.6%; Score 34.6; DB 12; Length 561;
Best Local Similarity 54.3%; Pred. No. 98;
Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

OY 51 TGCCTGCGCGCGCTGCTTCACAAATGGGCGCGCGGTATCATTAACGGCGCGGCA 110
DB 230 TGTCTCTGCGCGCGGTACCGGAAGACGGGAGATGACCGCGGTGACCGGACGACGCGGCA 171
OY 111 TAGTTCGGCGCGGACTATGATCAGCTGTACCGGTGTATCCATGGAATGGCACA 170
DB 170 TCTTGACAGCGGTGATCATGATCAGACACCTCATGAGTGTGATCTCACCAGCAGC 111
OY 171 TGCACCTGCA 179
DB 110 CGCTCTGCA 102

RESULT 15
BJ000263/c 640 bp mRNA linear EST 05-DEC-2001
LOCUS BJ000263 MF01SSA cDNA Oryzias latipes cDNA clone MF01SSA014A01 5',
DEFINITION mRNA sequence.
ACCESSION BJ000263
VERSION BJ000263.1 GI:17358166
KEYWORDS EST.
SOURCE Oryzias latipes (Japanese medaka)
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Percomorphi; Atherinomorpha;
Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
REFERENCE 1 (bases 1 to 640)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers

FEATURES
source
1..640
/organism="Oryzias latipes"
/mol_type="mRNA"
/strain="Hd-r"
/db_xref="taxon:8090"
/clone="MF01SSA014A01"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
/clone_lib="MF01SSA cDNA"

ORIGIN
Query Match 7.6%; Score 34.6; DB 12; Length 640;

Best Local Similarity 54.3%; Pred. No. 1e+02;
 Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY	51	TGCTCTGGCTGGCGCTCGTTCACATAGGGGGGGGGGATATCATACGGCGCGCGCA	110
Db	387	TGCTCTGGCGGGGCTAGGGAGAGGGGATGACGGGGTACCGCGGACGAGGCGA	328
QY	111	TAGTTCGGCCCGGACTATGATCAGTGTGTTACCGGTTGTTACCATGAATGGCACA	170
Db	327	TCTTGACGGCGTGTGATCATGATCAGCAGCTCCATCAGGTTGTGTTGATCTCACCGCAGC	268
QY	171	TGCACTGCA	179
Db	267	CGCTCTGCA	259

Search completed: March 16, 2004, 04:29:12
 Job time : 2235.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds
(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgc.....ccacgcctaacacgtactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: GenEmbl: *
2: gb_ba: *
3: gb_hcg: *
4: gb_in: *
5: gb_om: *
6: gb_ov: *
7: gb_pat: *
8: gb_ph: *
9: gb_pl: *
10: gb_pr: *
11: gb_ro: *
12: gb_sts: *
13: gb_sy: *
14: gb_un: *
15: gb_vi: *
16: gb_wa: *
17: em_fun: *
18: em_hum: *
19: em_in: *
20: em_mu: *
21: em_om: *
22: em_ov: *
23: em_ph: *
24: em_pl: *
25: em_ro: *
26: em_sts: *
27: em_un: *
28: em_vi: *
29: em_wa: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rnd: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_hcg_hum: *
39: em_hcg_mus: *
40: em_hcg_hum: *
41: em_hcg_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412.8	90.5	456	6	144909
2	412.8	90.5	2067	1	SEU43280
3	404.8	88.8	5103	1	STRAJ2301
4	404.8	88.8	22411	1	AE008749
5	403.2	88.4	254050	1	AL627269
6	403.2	88.4	301983	1	AE016840
7	397.4	87.1	1048	1	STAGRBA
8	288.6	63.3	361	6	144908
9	280	61.4	2889	1	CSP515700
10	238.4	52.3	4680	1	ECCSGABDG
11	238.4	52.3	10346	1	AE000205
12	238.4	52.3	15047	1	D90741
13	237.2	52.0	2920	1	CPR515701
14	236.8	51.9	456	6	AX814811
15	236.8	51.9	648	1	ECOCSCGAA
16	228.6	50.1	1711	1	AF275733
17	228.6	50.1	10190	1	AE005315
18	228.6	50.1	327773	1	AP002554
19	227	49.8	306358	1	AE016759
20	195.2	42.8	2883	1	ESAS15702
21	185.8	40.7	230	1	SEU53207
22	169.6	37.2	10370	1	AE015131
23	169.6	37.2	292504	1	AE016981
24	146.6	32.1	437	1	AF237726
25	93.2	20.4	19201	1	D90742
26	64.6	14.2	1212	1	EC0131756
27	48.2	10.6	78	6	AX814809
28	44.8	9.8	72	6	AX814798
29	42	9.2	2000	6	AX655393
30	36.2	7.9	110243	2	AC146945
31	36.2	7.9	148860	8	OSJN00093
32	35.8	7.9	298900	1	AP005937
33	35.6	7.8	151232	9	AC024086
34	35.6	7.8	185124	9	AC009358
35	35.6	7.8	186208	9	AC142338
36	35.4	7.8	2915	14	PAD459805
37	35.4	7.8	2929	14	PAD554049
38	35.4	7.8	3300	14	PAD311719
39	35.4	7.8	3315	14	PAU26221
40	35	7.7	5014	3	PEPFRNA
41	35	7.7	12641	1	AE014731
42	35	7.7	34980	6	AX492785
43	35	7.7	34980	6	AX535952
44	34.8	7.6	301727	1	AE016960
45	34.8	7.6	347625	1	BX248356

ALIGNMENTS

RESULT 1
LOCUS 144909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION 144909
VERSION 144909.1 GI:2469622
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinsen,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
Salmomella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

FEATURES
 source
 Location/Qualifiers
 1..456
 /organism="unknown"
 /mol_type="unassigned DNA"

Query Match 90.5%; Score 412.8; DB 6; Length 456;
 Best Local Similarity 94.1%; Pred. No. 6,6e-108;
 Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

ORIGIN

```

QY 1 ATGAAACTTTTAAAGTGACGATTCGACGAACTGAGTTCGACAGTCTGCGT 60
DB 1 ATGAAACTTTTAAAGTGACGATTCGACGAACTGAGTTCGACAGTCTGCGT 60
QY 61 GCGCTGCTTCCACATGAGGCGCGCGGTATCATTAACGCGCGCAATAGTTCGCG 120
DB 61 GCGCTGCTTCCACATGAGGCGCGCGGTATCATTAACGCGCGCAATAGTTCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCGTTCGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCGTTCGCA 180
QY 181 AGCGATGCGCGTAAATATGATCAGTACGCTTACCGGTGTGTACCATGAAATGACAT 240
DB 181 AGCGATGCGCGTAAATATGATCAGTACGCTTACCGGTGTGTACCATGAAATGACAT 240
QY 241 GCAGGCGCAGGCTGGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 GCAGGCGCAGGCTGGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCCACATGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 GCCACATGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AATPAAGCGCGCTGCTGATATCAGACCGCATGATGATGATGATGATGATGAT 420
DB 361 AATPAAGCGCGCTGCTGATATCAGACCGCATGATGATGATGATGATGATGAT 420
QY 421 GGTTTGGGACACACGCGCATACCGATTA 456
DB 421 GGTTTGGGACACACGCGCATACCGATTA 456

```

RESULT 2
 SEU43280 2067 bp DNA linear BCT 14-FEB-1996
 LOCUS
 DEFINITION Salmonella enteritidis agfBAC operon: fimbria-like protein
 precursor (agfB), thin aggregative fimbriae precursor (agfA), and
 AgfC (agfC) genes, complete cds.
 ACCESSION U43280
 VERSION U43280.1 GI:1184712
 KEYWORDS
 SOURCE
 ORGANISM
 Salmonella enteritidis
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.
 REFERENCES
 1 (bases 1 to 2067)
 Doran, J.L., Collinson, S.K., Burian, J., Sarlos, G., Todd, E.C.,
 Munro, C.K., Kay, C.M., Baner, P.A., Peterkin, P.I., and Kay, W.W.,
 DNA-based diagnostic tests for Salmonella species targeting agfA,
 the structural gene for thin, aggregative fimbriae
 J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
 JOURNAL MEDLINE
 PUBMED 8104955
 2 (bases 1 to 2067)
 Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and
 Kay, W.W.
 Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae
 J. Bacteriol. 178 (3), 662-667 (1996)
 JOURNAL MEDLINE
 PUBMED 850497
 3 (bases 1 to 2067)

AUTHORS
 TITLE
 JOURNAL
 SOURCE
 Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
 Direct Submission
 Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
 Microbiology, University of Victoria, P.O. Box 1055 Petch Bldg.,
 Victoria, BC V8W 3P6, Canada
 Location/Qualifiers
 1..2067
 /organism="Salmonella enteritidis"
 /mol_type="genomic DNA"
 /strain="27655-3b"
 /db_xref="taxon:592"
 /map="between puta and pyrC"
 571..576
 598..603
 696..1151
 /gene="agfB"
 /note="fimbria-like protein precursor"
 /codon_start=1
 /transl_table=11
 /product="AgfB precursor"
 /protein_id="AAC43598.1"
 /db_xref="GI:1184713"
 /translation="MKKLLPMLTILGAPGATATNDLARSFNRVNLSSFN
 SLSQAYGNSAIIQKSGNKANTITGYGTQTAAYVQKSHMAIRVQR"
 696..752
 /gene="agfB"
 753..1148
 /gene="agfB"
 /product="AgfB"
 /note="putative signal sequence of 21 amino acids;
 fimbria-like protein"
 1193..1648
 /gene="agfA"
 1193..1648
 /gene="agfA"
 /note="thin aggregative fimbria subunit precursor; major
 fimbrial subunit of thin aggregative fimbriae precursor"
 /codon_start=1
 /transl_table=11
 /product="AgfA fimbria precursor"
 /protein_id="AAC43599.1"
 /db_xref="GI:1184714"
 /translation="MKLLKVAAPFAIVGSGALAGVPMGGGNNHNGSGNSGPDST
 LSIYQGSANAALALQSDARKSRTITGSGYGADVGCGANSTIELTQNGFRNNAT
 IDQWNAKNSDITGVGGNNALVNTASDSSVMYRQVGFNNATANDY"
 1292..1354
 /gene="agfA"
 1352..1645
 /gene="agfA"
 /product="AgfA fimbria"
 /note="thin aggregative fimbria subunit; major fimbrial
 subunit of thin aggregative fimbriae"
 1667..1696
 1710..2036
 /gene="agfC"
 1710..2036
 /gene="agfC"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /product="AgfC"
 /protein_id="AAC43600.1"
 /db_xref="GI:1184715"
 /translation="MTLLALASNOITPTTQGDIVYIPIVTLNBPVCVOYILT
 SVRDSVGQSHQKTLSPANQPIEHSRLSVIISSEDSVKLIIVYSDQSLHLSQO
 WPSAQ"

ORIGIN

Query Match 90.5%; Score 412.8; DB 1; Length 2067;
 Best Local Similarity 94.1%; Pred. No. 7.2e-108;


```

gene
3344. 3376
/gene="csGBA"
-35_signal
3344. 3349
/gene="csGBA"
-10_signal
3371. 3376
/gene="csGBA"
gene
3469. 3924
/gene="csGB"
CDS
3469. 3924
/gene="csGB"
/function="nucleator"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CsGB protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/translacion="MKKLLFMWLTITGAPGATATATNDLASEYNFAVNEISKSEFN
QAALIGVGTNDNARVROBESKLSVTSQEGGNRAKVDQAGNPNFAYIEOTGNANDA
SISOSAYGNSAALITQKSGNKANTITQYGTOKTAVVQKSHMAIETVOR"
3469. 3931
/gene="csGB"
3532. 3921
/gene="csGB"
/product="CsGB protein"
/function="nucleator"
3966. 4421
/gene="csGA"
3966. 4421
/gene="csGA"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CsGA protein"
/protein_id="CA05317.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/translacion="MKLKVAAPFAITVSGSALAGVYPOMGGGNGNNGSGSPDST
LSIYOGSANAALALOSDARKSETTITGSGADVGQAGNSETTELTONGRNNAT
IDOMNAKNSDITVGQYGGNNALVNOIRASDSVWVQVGGNNATANQY"
3966. 4025
/gene="csGA"
4026. 4418
/gene="csGA"
/product="CsGA protein"
/function="fiber subunit"
4483. 4809
/gene="csGC"
4483. 4809
/gene="csGC"
/codon_start=1
/transl_table=11
/product="CsGC protein"
/protein_id="CA05318.1"
/db_xref="GI:2739240"
/db_xref="SWISS-PROT:P55227"
/translacion="MHTLLIALNSQITFTTQGGDIYTVIPQVTLNPPCYCOVOIT
SVNDVGQGGHTQOKTSLIPANQPIELSLSTVNISSESVKTIYTVSDGSLHSQO
WPSAQ"
4483. 4533
/gene="csGC"
4534. 4806
/gene="csGC"
/product="CsGC"
4940. 5103
/gene="ORF103"
4940. 5103
/gene="ORF103"
/codon_start=1

```

```

ORIGIN
Query Match      88.8%; Score 404.8; DB 1; Length 5103;
Best Local Similarity 93.0%; Pred. No. 1.6e-105;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

/transl_table=1
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:O54295"
/translacion="MSVYKKRNPATIGLCTCAFTIHSAGQQTVOGSHIFRGAIVEPL
CDISTHMENT"

RESULT 4
AE008749      22411 bp      DNA      linear      BCT 23-APR-2003
LOCUS
DEFINITION
Salmonella typhimurium LT2, section 53 of 220 of the complete
genome.
ACCESSION
AE008749 AE006468
VERSION
AE008749.1 GI:16419641
KEYWORDS
SOURCE
ORGANISM
Salmonella typhimurium LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 22411)
REFERENCE
McLelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
Latreille,P., Courtney,L., Porwolik,S., Ali,J., Dante,M., Du,F.,
Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Flores,L., Miller,W.,
Stonking,T., Nhan,M., Waterston,R. and Wilson,R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL
Nature 413 (6858), 852-856 (2001)
MEDLINE
21534948
PUBMED
11677609
REFERENCE
1
2
The Salmonella typhimurium Genome Sequencing Project
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of

```

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI33283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and Genemark; <http://opal.biology.gatech.edu/Genemark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; [http://ecocyc.org/ecocyc/](http://ecocyc.org/ecocyc/PangeaSystems.com/ecocyc/)

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB;

http://kinich.cftn.unam.mx:8850/db/reguondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

FEATURES

SOURCE

Location/Qualifiers

1..22411

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

/gene="phoH"

/note="synonym: STM1126"

434..439

/gene="phoH"

/note="putative RBS for phoH; ReguonDB:STMS1H001398"

454..1308

/gene="phoH"

/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"

/codon_start=1

/transl_table=1

/product="phoB-dependent ATP-binding pho regulon component"

/protein_id="AAL20057.1"

/db_xref="GI:16419642"

/translation="MCANRVQENAGDVYQLKYLISMGROKAVIKARREARVLRDS RSHKQREESVTSLVQMGVEAIGWARSRPTSPVKANEQAHLNADSKQIFAT

GEACGKTWISAAGAALIKHDEVERITVTRPVLOADEDLGLPDGIDAEKAPFRRPV YDVLKRLGASFMQYCLRPETGKVIAPFAPRGTPENAVVIIIDEAQNTAAAMKMP

LTRRGNTVTVNGDITQCDLPRGVRSLDALERFEDENVGIVHFKDKCSALC ORTHAYS"

complement(1414..2303)

/gene="STM1127"

complement(1414..2295)

/gene="STM1127"

/note="similar to E. coli orf, hypothetical protein (AAC7480.1); Blastp hit to AAC7480.1 (285 aa), 23%

identity in aa 15 - 264"

/codon_start=1

/transl_table=1

/product="putative transcriptional regulator"

/protein_id="AAL20058.1"

/db_xref="GI:16419643"

/translation="MEQPRLPKPKGKILDTLGMOKSITRASORIAQYIAPRQVTO SSIDLSRETQAGATVTRFCRTLGKGFQFQDKDLAIELATTSDDSSPLDLEVSE

SDDAHAIQIKLQNTISNVLTSLNLIDQVLTGVADLRHCHSVYIFVGGSGITLAD MKHKLMRIGLRGDAVSNHFMVQATLLKADVAMGVSHSGTSPETVYSLRLAQAG

RBS

TTVAITHNIGSLCEADFCILINGRQMGQDSIGTKAQLFVPDLTYTLVQSSPE OARSKARTNALMTK"

complement(2258..2303)

/gene="STM1127"

/note="putative RBS for STM1127; ReguonDB:STMS1H001399"

complement(2580..4088)

/gene="STM1128"

complement(2580..4076)

/gene="STM1128"

/note="similar to E. coli putative cotransporter (AAC76102.1); Blastp hit to AAC76102.1 (571 aa), 22%

identity in aa 7 - 478"

/codon_start=1

/transl_table=1

/product="putative sodium/glucose cotransporter"

/protein_id="AAL20059.1"

/db_xref="GI:16419644"

/translation="MTHSGFIVNVYLVFGYLLAMLVGVYFSRROKTADYFRGGR VPGWAGVSVATLSTITPMSIPAKAFSTSDWTFIIGQYLAAILPLVYFYIPERK

LKYSABEYLARDVRCRLPASPMLPHIGRAIITFPLVLRPAIDPVLIVL LYSVCITITVMGSGIEGVITWDVIOGLLSGAILIFVLCVKQGGIDEIFVTQQA

DKFPPAQFHWSTESTVPLVIMIGFLPANIQQFASODVORVYVTSIBETKTLT NAKLVAVIYPVFPFAGSALFVYQOHPQLPAGNTGGLPLFVTEMPVGIAGIIA

AIIPAASISISSNISCSFNSDIYORLSHKRTPEKMKIAKLIVLWAGISPAAS VMLVMADESEIMDAFNSLIGMGSPMTGIFFPFRANSGAVIGIIIVITVGA

RYATDINPFYGVIGSLSVITSGVIFAPLPAPAPPLIDEXPERKTLT"

complement(4083..4088)

/gene="STM1128"

/note="putative RBS for STM1128; ReguonDB:STMS1H001400"

complement(4413..5099)

/gene="STM1129"

complement(4413..5093)

/gene="STM1129"

/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -

226"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20060.1"

/db_xref="GI:16419645"

/translation="MSLARLEQSHVHENGGLIVSCOPVPSPMDEKPELVAAQMAAS AGAVAVLEGIEGRTVPHLSVPIIGIRKDLGSPRTPIQDPAALAAQADITP

AFDASFRSPVDISLIRIRHGLMLMADCSYTNBEGISCHQKTEFRTGLSGTGP ITPEPLWATQLSHAGCRVLAEGRYNTPALANABHGAATVGSATIRIEHTQ

WFSHAVKR"

complement(5094..5099)

/gene="STM1129"

/note="putative RBS for STM1129; ReguonDB:STMS1H001401"

5599..6759

/gene="STM1130"

5599..6759

/gene="STM1130"

/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%

identity in aa 40 - 403"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20061.1"

/db_xref="GI:16419646"

/translation="MGMQKPKFKMTLALCSVAITTSQYATLTDIPPLKNGTG AIDNNGVIVGLGAGTSWKIDLKQKHMERIKSPGGAQREOSVSPFLDELIVRG

GVGKNSESPLOVQVSDYKYSPVKNQKQWERTISPVGLTGTGKMETVWLITGVN EHIIDPKYFIDIAAADADESKNKVIVYFNKPADYFNNKIVFYLYNAKENTKAGEL

PGATGAGSSVMGNPFMLINGELKPGRTDVIIRAMMDKQLFTLKNLSQLPSPGEO QOEGLAGFSGYSHVILVGGGANFPKAKQNTYNGKYSHEGINKKRDVYGLINCH

QOYNGKKQKQPLGVISYSGDEVILIGENAKGPVSVTSFTKRDGNLTK"

6793..7497

/gene="STM1131"

6793..6798

/gene="STM1131"

/note="putative RBS for STM1131; ReguonDB:STMS1H001402"

CDS

6805..7497
 /gene="STM1131"
 /note="similar to E. coli orf, hypothetical protein
 (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 268
 identity in aa 4 - 239"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAL20062.1"
 /db_xref="GI:16419647"
 /translation="MKINKYLLGMVFSFSSYLQANTLDYRHYADRTINQRIAI
 IEKPNIGIFVDASVKSQGVGDQDKLSDLVANALIEGVSINIKYDVFVQPF
 FESGPDITIKYKRGQYNDPSGYTMGRYRYPDARTANYSDEKTRDYIGV
 DELKLENYFMWDSQIKFNDKNTNYEHVVALMKLNSFTPYVEGVNAVNNITDR
 OTRVVGLOVHF"
 7768..9060
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
 7768..7773
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
 7780..9060
 /gene="STM1132"
 /note="similar to E. coli putative transport protein
 (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 574
 identity in aa 25 - 418"
 /codon_start=1
 /transl_table=11
 /product="putative sugar transport protein"

RBS

Query Match 88.8%; Score 404.8; DB 1; Length 22411;
 Best Local Similarity 93.0%; Pred. No. 17e-105;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

CDS

1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTCTTGAGAGCTCTGGCT 60
 17769 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTCTTGAGAGCTCTGGCT 17828
 61 GGGCTGCTTCCAAATGGGCGCGCGGTATCAATACGGCGGCGCAATAGTTCGGC 120
 17829 GGGCTGCTTCCAAATGGGCGCGCGGTATCAATACGGCGGCGCAATAGTTCGGC 17888
 121 CCGGACTCAACGTTATGAGATGATGAGTTCGGCTACGCTGCGCTGCTGCTGCA 180
 17889 CCGGACTCAACGTTATGAGATGATGAGTTCGGCTACGCTGCGCTGCTGCTGCA 17948
 181 ACCGATGCCCGTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 17949 ACCGATGCCCGTAATGATGATGATGATGATGATGATGATGATGATGATGAT 18008
 241 GCAAGCCAGAGGTGGCGATATAGTATGATGATGATGATGATGATGATGATGAT 300
 18009 GTAGGCGCAGAGGTGGCGATATAGTATGATGATGATGATGATGATGATGAT 18068
 301 GCCACATCGACCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 18069 GCCACATCGACCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 18128
 361 AATAACCGCGGCTGTTTATCAGACCGCATGATGATGATGATGATGATGATGAT 420
 18129 AATAACCGCGGCTGTTTATCAGACCGCATGATGATGATGATGATGATGAT 18188
 421 GGTTCGCAACACCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 456
 18189 GGTTCGCAACACCGCATGATGATGATGATGATGATGATGATGATGATGAT 18224

RESULT 5

AL627269 254050 bp DNA linear BCT 04-JUL-2003
 LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
 DEFINITION complete chromosome, segment 5/20.
 ACCESION AL627269 AL513382
 VERSION AL627269.1 GI:16502231

KEYWORDS

SOURCE Salmonella enterica subsp. enterica serovar Typhi

ORGANISM Salmonella enterica subsp. enterica serovar Typhi

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

REFERENCE

1 (bases 1 to 254050)

Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,

Wain, J., Churcher, C., Mungall, K. L., Bentley, S. D., Holden, M. T. G.,

Seshadri, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T.,

Comerton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N.,

Farrar, J., Feltham, T., Hamlin, N., Haque, A., Hien, T. T., Holroyd, S.,

Jagels, K., Krogh, A., Larsen, T. S., Leather, S., Moule, S., O'Gaora, P.,

Perry, C., Quail, A., Rutherford, K., Simmonds, M., Skelton, J.,

Stevens, K., Whitehead, S. and Barrall, B. G.

Complete genome sequence of a multiple drug resistant Salmonella

enterica serovar Typhi CT18

Nature 413 (6858), 848-852 (2001)

TITLE

Journal

Medline

PubMed

Reference

Authors

Journal

Comment

Notes

Details of S. typhi sequencing at the Sanger Centre are available

on the World Wide Web.

(URL: http://www.sanger.ac.uk/Projects/S_typhi/).

Location/Qualifiers

1..254050

/organism="Salmonella enterica subsp. enterica serovar

Typhi"

/mol_type="genomic DNA"

/strain="CT18"

/db_xref="taxon:90370"

181..2793

/gene="STY1078"

/note="synonym: pepn"

181..2793

/gene="STY1078"

/note="Orthologue of E. coli pepN (AMPN ECOLI); Pasta hit

to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"

/codon_start=1

/product="aminopeptidase N"

/protein_id="CAD08183.1"

/db_xref="GI:16502232"

/db_xref="GCA:Q82770"

/db_xref="SPTREMBL:Q82770"

/translation="MTQPOAKYRHDYRVDYQITDIETFDLDEKTVTAISQAVR

HGAPADLRIDGDLTVLSIHVNDAPWYKKEEALISLDPREFTRINELISPA

NTALBGLYOSDALCTOCBAEGFRHTWVLPDVLARPTTKITDKSKIFPLISPA

RVAQGLENGRLHWVQODPEPRCYLPFLVAGDPVLDPTTSGRVALDELVDG

NIDRAVAMTSLKSMKDEPRRGLEYDIDITVAIVDFEMGAMENGLNIFNSKY

LAVDTATDKDVIDIERTIGHEYFNWNGINRQCDLWQLSKGLTVFRQESSDL

TLLGERNFQKQQLYFERRDGSATCDPVOAMEASVNDVLTVEKGAETPLRMH

KODVNEDEQYTLTISQRTPAVADQAEQPHIFPALEIYNENGVNPLQKGGPVNA

VANVQAEQTFEDNVVFOVPALCEFSAVKLEVKMSDQGLFLMRHANSRND

IAELFVADIPDLIAQVREALRRLTAAELADEFLAIYVANHIDEYRVHGDIGKRLN

ACRFLFTGSETEIANTVYSKQYRDANNNTDIALALSAVAQAOLPCRDITMOEYDQK

ADGIVMDKFILOSTSPAEVLETVSRSLKRSFSMSNPRIISLIGAFGSPAPH

ADGSGYQPLVEMLTDLNSNPNVQASRLIPLIRLKYRDKRQEKMALEQKGLN

LSGDIKITYALA"

532..1332

/gene="STY1078"

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

family M1, score 245.10, E-value 1.3e-84"

1060..1089

/gene="STY1078"

misc_feature

misc_feature

gene /note="PS00142 Neutral zinc metalloproteinases, zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:082759"
/db_xref="SWISS-PROT:082759"
/translation="MYPPVRRKALFQDLPERAHEFTFOQLRITGTPLLEALVROKVP
KPVCMGLTFKRLPLGLAGLDKDECDALGAMGFGSLIGTVPRPOPNDKPLFR
LVDAEGLINRMGFNNLGVNDLVENYKKAHPDGLIGINIKNDTPVENGRDYLIME
KVYAVAGYIAINISSPNTPLGRTIYQGDALDILTKKNDLQVTHKKVPAVAKI
APLCEBELIOVADSLRHNDGVIATTTTIDRSIVQMKRCQQTGGLSGRPLQKST
EITRLSLKGLPLIGVGIDSVIAREKIALAGATLVQISGTFKGPLKEIYV
HI"
misc_feature 3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODHase, dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBM_ECOLI; Faeta hit to YCBM_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTRMBL:082758"
/translation="MMRIKPDNDNRWYDEEDHDMMLDLANGMLFRSFRSKMLTPDA
FCPTGFCVDNALYFSPFEKCRDELTKGKRAELVNLALVIRYLKQMPKSMHFAVA
GEMWTPNGDAASVLSLDTAEOVNLVVEPBNALCLLAPGVVIAGRTQLDALK
IMNRLKRVQVCHSFSLEQAV"
CDS complement(4716..5825)
/gene="STY1083"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:082757"
/db_xref="SPTRMBL:082757"
/translation="MATLSRLFIHPVSMRGIGLTHALDISGLAFRIEMVTSDEGT
FITAROPQWRFTPSPLHDGLHTAPGSSALVRFIDFTQDAPTEWGNHFAARA
PAINOMLSGFSRDVLCWGPQPLTRKVKHNAVPIGADGYLLTNESLSLDLOO
RCPAGVQMEQFSPNLVYGVAAWEDSKVRIQGVIPDYKPCSRICFTTYSPEKQ
KHSGCPALTIQAFRTADNDGVDFGQNLIRNSGVRVSGVEVILATAPKAYGATT
LDISTVPEKHPDGVTTIDMGQITFCVNNQVLLLEQLQKQIRIYPSGAGICGCRIR
LLEGEVPLKKSAGSDGTILISCSVPKTAIRLEN"
misc_feature complement(4737..4955)
/gene="STY1081"

gene /note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 5.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein h1016 SW:YCBY_HAEIN (P44524; P43945) Faeta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:082756"
/db_xref="SPTRMBL:082756"
/translation="MNSLPASTARGLKIELEKAGVCCOVQGVHPOGTRLI
YQSLMWSRLSRILIPMGCKVTSDDLTVRGVQALNTEIFNPATRAVHPSGLNDTI
RNSQYGMKVKYDAIVDAFTKRLPRPVNDESPDLRINWLNKETAIALDLSGDLH
LRGVRDRTGLAPIKETLAAIIVMSGMPGLDPMCGSGTLLIEAMAWATDRAPIGL
HRGHMFGSGWAQDEAIWQSVKAEQTRARKGLAEYSHFYSGSDARVERARSNNR
RAGIGELITREVDVQALSNPLPKPGVTVSNPVGRELDSPDALTAHSLGRTMK
NORGAMLSIFSASPDLSIQLRADQFPAKNGPLDVCQKNTIATTTADSKPATYA
EDIANRRLKRLKLEKARQEGIECVLYADIPREVVAADRYDMVIOEYAPKTV
DAQKARQRLDIIAATLSVGIIPNKLVLTREKQKNOYQKXSEGEVLESEVNA
RLWNLTDYDGLFLDRIARMLGMSKGLVLFNFSYGSASVAGAGGAGATT
VMSRITYLEWAERMLRLNGSGRAHRLIADCGMLSEANQDFLIFIDPTFNSNR
MESFPDYQRHVLMKDLKRLKRGITIMSNKRGFRMDLBLEGLAEGLTAQETTKT
LSPPFARNROIHNCWILRAA"
misc_feature 5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Faeta hit to YJUK_ECOLI (554 aa), 34% identity in 524 aa overlap"
CDS 8045..9952
gene 8045..9952
Query Match 88.4%; Score 403.2; DB 1; Length 254050;
Best Local Similarity 92.8%; Pred. No. 5,7e-105;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGACGATTCGGACGATCGTAGTTCTGACAGTCTGGCT 60
DB 88906 ATGAACCTTTAAAGTGGACGATTCGGACGATCGTAGTTCTGACAGTCTGGCT 88965
QY 61 GGCCTCGTTCACATATGGGCGGCGGCGGTATCATATACGGCGGCAATAGTCCGGC 120
DB 88966 GGCCTCGTTCACATATGGGCGGCGGCGGTATCATATACGGCGGCAATAGTCCGGC 89025
QY 121 CCGGACTGAACTTGACATTTATCAGTACGCTTCGGCTAACCGTGGCTTCTTGCAA 180
DB 89026 CCGGACTGAACTTGACATTTATCAGTACGCTTCGGCTAACCGTGGCTTCTTGCAA 89085
QY 181 AGGATGCCGTAATATGATGATCGCTGTATACCGGTGTGTTACCATGAATAGGCAT 240
DB 89086 AGGATGCCGTAATATGATGATCGCTGTATACCGGTGTGTTACCATGAATAGGCAT 89145
QY 241 GCAAGCCAGGAGTGGATTAATAGTACTATTGAATGAACTGAATGAGTGTTCAGAAATAT 300
DB 89146 GTAGCCAGGAGTGGATTAATAGTACTATTGAATGAACTGAATGAGTGTTCAGAAATAT 89205

Oy	301	GCACCATGACCACTGGAGCGTAAACTCCGATATTACTGTGCCGCAATAAGCGCGT	360
Dd	89206	GCCACCATCAGCACTGGAGCGTAAACTCCGATATTACTGTGCCGCAATAAGCGCGG	89265
Oy	361	AATAAGCGCGCGTGTTAATCAGCCGATCTGATTCAGCGGTAATGCGTCAGGTT	420
Dd	89266	AATAAGCGCGCGTGTTAATCAGCCGATCTGATTCAGCGGTAATGCGTCAGGTT	89325
Oy	421	GTTTTGGCAACAACGCGCGCTAACGATATTAA	456
Dd	89326	GTTTTGGCAACAACGCGCGCTAACGATATTAA	89361
Db	89326	GTTTTGGCAACAACGCGCGCTAACGATATTAA	89361
RESULT 6			
LOCUS	AE016840/c	301983 bp DNA linear BCT 21-MAR-2003	
DEFINITION		Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of	
ACCESSION		AE016840 AE014613	
VERSION		AE016840.1 GI:29137797	
KEYWORDS			
SOURCE			
ORGANISM		Salmonella enterica subsp. enterica serovar Typhi Ty2	
REFERENCE		Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	
AUTHORS		Enterobacteriaceae; Salmonella.	
TITLE		1 (bases 1 to 301983)	
JOURNAL		Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,	
MEDLINE		Burland, V., Kodoyiamni, V., Schwartz, D.C. and Blattner, F.R.	
PUBMED		Comparative Genomics of Salmonella enterica Serovar Typhi Strains	
AUTHORS		J. Bacteriol. 185 (7), 2330-2337 (2003)	
TITLE		2 (bases 1 to 301983)	
JOURNAL		Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,	
MEDLINE		Burland, V., Kodoyiamni, V., Schwartz, D.C. and Blattner, F.R.	
PUBMED		Direct Submissions	
AUTHORS		Submitted (25-SEP-2002) Laboratory of Genetics, University of	
TITLE		Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA	
JOURNAL		Location/Qualifiers	
MEDLINE		1. 301983	
PUBMED		/organism="Salmonella enterica subsp. enterica serovar	
AUTHORS		Typhi Ty2"	
TITLE		/mol_type="genomic DNA"	
JOURNAL		/strain="Ty2"	
MEDLINE		/serovar="Typhi"	
PUBMED		/sub_species="enterica"	
AUTHORS		/db_xref="taxon:209261"	
TITLE		complement(164..622)	
JOURNAL		/gene="tnpa"	
MEDLINE		/locus_tag="t1735"	
PUBMED		complement(164..622)	
AUTHORS		/gene="tnpa"	
TITLE		/locus_tag="t1735"	
JOURNAL		/note="corresponds to STY1225 from Accession AL513382:	
MEDLINE		Salmonella typhi CT18"	
PUBMED		/codon_start=1	
AUTHORS		/transl_table=11	
TITLE		/product="transposase for insertion sequence element	
JOURNAL		IS200"	
MEDLINE		/protein_id="AA069359.1"	
PUBMED		/db_xref="GI:29137798"	
AUTHORS		/translation="MGDEKSLATRNCKYHIVAPKYRRAFGGEKRASITIRKI	
TITLE		CENNVNIIIEACCADHIHLMLIETPKMSVSPMVKYSIMTYEOPGLKKRYRN	
JOURNAL		REWRGRVYVDYTKNTAKTIQDIKHOLEDMGRLSIPIGSPFTGRK"	
MEDLINE		complement(844..1797)	
PUBMED		/gene="figL"	
AUTHORS		/locus_tag="t1736"	
TITLE		complement(844..1797)	
JOURNAL		/gene="figL"	
MEDLINE		/locus_tag="t1736"	
PUBMED		/note="corresponds to STY1223 from Accession AL513382:"	
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
AUTHORS			
TITLE			


```
gene
/RAALNAGATPMDILSLOSAGCLRAKLEII"
/complement(5597..6295)
/gene="flgH"
/locus_tag="t1740"
/complement(5597..6295)
/gene="flgH"
/locus_tag="t1740"
/feature="corresponds to STY1219 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AAO69364.1"
/db_xref="GI:29137803"
/translation="MQRALAHYPMALMVAATLTCGAMIPAKPLVQCATTTQPIPGPV
PVANSITQSAQPIVGTQPLFEDRPNNIGTTLTVQENVASKSSANASRDGKT
SPGFTVPRVYLOGLEGNRADMEASGNSFGKGGANASNTFSGTLTVTVQVLANGN
LHVVEKQIAINQCTEFIRFSGVNPRTISGNSVPTQVADARIEVYNGYINBAON
MGIORFPLNLSPM"
/complement(6350..7132)
/gene="flgG"
/locus_tag="t1741"
/complement(6350..7132)
/gene="flgG"
/locus_tag="t1741"
/feature="distal rod protein: corresponds to STY1218 from
Accession AU513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flgG"
/protein_id="AAO69365.1"
/db_xref="GI:29137804"
/translation="MISLSLWAKTGDLDAQOTNMVIANNLAVSTNGSKRQRAVEDL
LYQTRQGAQSEDTLPSSQIGTGVPRATERLHQSQNLSTNNKSDVAIKQGF
FOVNLPGDTSAVTRDGSQVQVNDQGLVAGQVQVPAITIPANALSTIGRDGVSVT
QOQGAAPVQVQGLNLTTFMNDTGLSEIENLTETOSGAPNESTPGLNGAGLLYQGY
VETSNVAEELVMMIOVRAYEINSKAVSTTDQMLKRLQL"
/complement(7146..7901)
/gene="flgF"
/locus_tag="t1742"
/complement(7146..7901)
/gene="flgF"
/locus_tag="t1742"
/feature="corresponds to STY1217 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AAO69366.1"
/db_xref="GI:29137805"
/translation="MDHALITRMGAASQTLNQAVTASNLANASTPGRADLNLRAV
FVDSLATKRLVTRASTPGADMTQQLDYSRPLDVALQDQGMVLVQADAEQTRN
GNIVGPTGQGLTIQGHPIVIGGSEITVEGESEITADPISALNPGDPPTVAIPVR
LKLVAEGNEVQRSDDLGFLRLAEQKAEKRGAVLADPSIRIMSGVLSESNVKEAMT
DMANARFEMQMKVITSVDNEGRANQLLMS"
/complement(7922..9133)
/gene="flgE"
/locus_tag="t1743"
/complement(7922..9133)
/gene="flgE"
/locus_tag="t1743"
/feature="corresponds to STY1216 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flgE"
/protein_id="AAO69367.1"
/db_xref="GI:29137806"
/translation="MSPSOAVSGLNAAATNLDTGNNTANSATYGFKSGTSPADMA
GSKVGLGVKAVAGITQDFDGTNTTGRGLDVAISQNGFPRLVDSNGSVFYSRNQDFKL
```

Query Match 88.4% Score 403.2? DB 1; Length 301983;
Best Local Similarity 92.8% Pred. No. 5.8e-105;

```
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGATTGTTTGGCAGTCTGGCT 60
DB 37310 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGATTGTTTGGCAGTCTGGCT 37251
QY 61 GGGCGTCTTCCACAAATGGGGGGGGGGGTTAATCAATACGGCGGCGGCAATAGTTCCGGC 120
DB 37250 GGGCGTCTTCCACAAATGGGGGGGGGGGTTAATCAATACGGCGGCGGCAATAGTTCCGGC 37191
QY 121 CCGACTCAACGTTAGCATTTATCAGTACGGTTCCGTTACGTTCCGTTCTCTGCA 180
DB 37190 CCGACTCAACGTTAGCATTTATCAGTACGGTTCCGTTACGTTCCGTTCTCTGCA 37131
QY 181 AGCGATGCGGCTTAATATGATCAGCTGGTTACCGGTGTTTACCATGAATGCGCAT 240
DB 37130 AGCGATGCGGCTTAATATGATCAGCTGGTTACCGGTGTTTACCATGAATGCGCAT 37071
QY 241 GCAGGCCAGGGTGGGATTAATAGTACTATTGAATGACTGCAAGTGGTTTCAAAATTAAT 300
DB 37070 GCAGGCCAGGGTGGGATTAATAGTACTATTGAATGACTGCAAGTGGTTTCAAAATTAAT 37011
QY 301 GCCACATCGACCACTGGAACGCTAAACTCCGATATTACTGTCGGCCATACGGCGGT 360
DB 37010 GCCACATCGACCACTGGAACGCTAAACTCCGATATTACTGTCGGCCATACGGCGGT 36951
QY 361 AATAACGCGCGCTGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGTT 420
DB 36950 AATAACGCGCGCTGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGTT 36891
QY 421 GGTTCGCAACACGCGCAGCGCTTAACAGTATTA 456
DB 36890 GGTTCGCAACACGCGCAGCGCTTAACAGTATTA 36855

RESULT 7
STRAPBA 1048 bp DNA linear BCT 26-JAN-1998
LOCUS Salmonella typhimurium agfB and agfA genes.
DEFINITION Salmonella typhimurium agfB and agfA genes.
VERSION AJ000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
Infect. Immun. 65 (12), 5320-5325 (1997)
JOURNAL
MEDLINE 98053981
PUBMED 9393832
REFERENCE
2 (bases 1 to 1048)
AUTHORS Sukupolvi,S.S.
TITLE Direct Submision
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinamyllynkatu, 20520, FINLAND
FEATURES
source
1..1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="AgfB protein"
/protein_id="CAA04150.1"
```

/db_xref="GI:2275120"
 /db_xref="SPTREMBL:O33801"
 /translation="MKKLLPMLTLLIIEPIATATNYDLARSEYNAVELSKSFNO
 AALIGQVGTNSARVHRSKLLSVISOERGNNRAKADQAGNYPAYIEQTNADA
 SISOAYNSAISAIIOKSGNKNKITGYQTKAVVVOQKSHAIQANITQYQTO
 KTAVVVOKSHMAIRVTR"
 593..1048
 /gene="agfa"
 593..1048
 /gene="agfa"
 /codon_start=1
 /transl_table=1
 /product="Agfa protein"
 /protein_id="CA04151.1"
 /db_xref="GI:2275121"
 /db_xref="SPTREMBL:O33802"
 /translation="MKLLKVAFAIIVSGSAVAVVPOWGGGNNHNGNSGPDST
 LSTIYQSSANAIALOSDARKSEITITQSYGAGADVGGAGNISTELTONFRNNAT
 IDOWNANSDITVGYGGNNMALVNOTASDSVMVQVGFNNAPANGN"
 ORIGIN

Query Match 87.1%; Score 397.4; DB 1; Length 1048;
 Best Local Similarity 92.1%; Pred. No. 1.9e-103;
 Matches 419; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGCAGCATTGCGACATGCTGTTCTGGCAGTCTGCTGCT 60
 DB 593 ATGAACCTTTAAAGTGGCAGCATTGCGACATGCTGTTCTGGCAGTCTGCTGCT 652
 QY 61 GGGCTGTTTCAACATGGGGCGGGCGGCTTAATCAATAGCGCGCAATAGTTCCGGC 120
 DB 653 GGGCTGTTTCAACATGGGGCGGGCGGCTTAATCAATAGCGCGCAATAGTTCCGGC 712
 QY 121 CCGGACTCAACGTTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCGCTGCTGCTGAA 180
 DB 713 CCGGACTTCCAGCTTGAGCATTTATCAGTACGGTCCGCTAACGCTGCGCTGCTGCTGAA 772
 QY 181 AGCGATGCCGCTAAATATGATCAGCTGCTTACCGCTGTTTACCATGAATAGCAAT 240
 DB 773 AGCGATGCCGCTAAATGTAATCTGAAACGACATTAACCCAGACGGCTTAATGTAACGCGCGCAT 832
 QY 241 GCAGGCGCAGGTCGCGATTAATGTAATTTGTAATGTAATGTAATGTAATGTAATGTAAT 300
 DB 833 GTAGGCGCAGGTCGCGATTAATGTAATTTGTAATGTAATGTAATGTAATGTAATGTAAT 892
 QY 301 GCCACCATGACCATGAGTGAACCTTAATTTACTGTCGCGCAATACGCGCGT 360
 DB 893 GCCACCATGACCATGAGTGAACCTTAATTTACTGTCGCGCAATACGCGCGT 952
 QY 361 AATAACGCCGCGCTGTTAATCAGACGCGCATCTGATTCGACGCTAATGTCGCTCAGGTT 420
 DB 953 AATAACGCCGCGCTGTTAATCAGACGCGCATCTGATTCGACGCTAATGTCGCTCAGGTT 1012
 QY 421 GCTTTGGCAACACGCCGCGCTTAACCATGATTA 455
 DB 1013 GCTTTGGCAACACGCCGCGCTTAACCATGATTA 1047

RESULT 8
 LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
 DEFINITION Sequence 56 from patent US 5635617.
 ACCESSION 144908
 VERSION 144908.1 GI:2469621
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 361)
 AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen, and Clourhier,S.C.
 TITLE Methods and compositions comprising the agfa gene for detection of
 JOURNAL Patent: US 5635617-A 56 03-JUN-1997;
 Salmonella

FEATURES
 source Location/Qualifiers
 1..361
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN

Query Match 63.3%; Score 288.6; DB 6; Length 361;
 Best Local Similarity 91.3%; Pred. No. 4.5e-72;
 Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 64 GTCGTTCCACATGGGGCGGGCGGCTTAATCATATACGCGCGCAATAGTTCCGGCCG 123
 DB 1 GTCGTTCCACATGGGGCGGGCGGCTTAATCATATACGCGCGCAATAGTTCCGGCCG 60
 QY 124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTTAACGCTGCGCTTCTGCAAAAGC 183
 DB 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTTAACGCTGCGCTTCTGCAAAAGC 120
 QY 184 GATGCCGCTTAATATGATCAGCTGTTACCCGCTGTTTACCCATGAATGGCACATCA 243
 DB 121 GATGCCGCTTAATATGATCAGCTGTTACCCGCTGTTTACCCATGAATGGCACATCA 180
 QY 244 GGGCAGGTCGCGATTAATGTAATTTGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 303
 DB 181 GGGCAGGTCGCGATTAATGTAATTTGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
 QY 304 ACCATGACACGATGGAACGCTAAATCCGATTTATCTGTCGCGCATAGCGCGTAA 363
 DB 241 ACCATGACACGATGGAACGCTAAATCCGATTTATCTGTCGCGCATAGCGCGTAA 300
 QY 364 AACGCGCGCTGTTAATCAACCGCATCTGATTC 398
 DB 301 AACGCGCGCTGTTAATCAACCGCATCTGATTC 335

RESULT 9
 LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
 DEFINITION Citrobacter sp. Fec2 csb gene, csbA gene and csbD gene.
 ACCESSION AJ515700
 VERSION AJ515700.1 GI:31790491
 KEYWORDS csbA gene; csbB gene; csbD gene; curlin-csbA protein; nucleation
 component of curlin monomers; regulatory protein.
 SOURCE Citrobacter sp. Fec2
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Citrobacter.
 REFERENCE 1
 AUTHORS Zogaj,X., Bokranz,W., Nimetz,M., and Romling,U.
 TITLE Production of Cellulose and Curli Fimbriae by Members of the Family
 JOURNAL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
 INFECTION, Immun. 72 (7), 4151-4158 (2003)
 AUTHORS Romling,U.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
 CENTER, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES
 source Location/Qualifiers
 1..2889
 /organism="Citrobacter sp. Fec2"
 /mol_type="genomic DNA"
 /isolate="Fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:213763"
 /country="Germany"
 /complement(212..862)
 /gene="csbD"
 /complement(212..862)
 /gene="csbD"
 /function="necessary for transcription of the csbA
 operon"
 /codon_start=1

gene
 CDS
 1 (bases 1 to 2889)
 /organism="Citrobacter sp. Fec2"
 /mol_type="genomic DNA"
 /isolate="Fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:213763"
 /country="Germany"
 /complement(212..862)
 /gene="csbD"
 /complement(212..862)
 /gene="csbD"
 /function="necessary for transcription of the csbA
 operon"
 /codon_start=1

```
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MFNEFHSIHGHTLLITKPSIQATALLQHKQSIALSGKLNIIQ
RSLDDISSCIVLDMMEDKLIHYMDNLSRKNNIKTLLITPDYPRDIENMP
HINQFVYADDEERYNGLQILRECFYSGKSLASLTTHSGNRYNSSEFALLTHRE
KEILNKIRIGASNIENIARSLFISENTVTHLYNLFKIAVNRQAVSWANDNLR"
1623..2078
/gene="csgb"
1623..2078
/gene="csgb"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKNKLLFMMLTVLGAPGIASNTSYDLAHEXNPAVNEISKSPN
OAILIGVGTNNSAKRQEGSKLSVYSQEGSGRPARAKDQGANFATIAQSHSND
SISQSNVNTMTIIOKSGNKNANTQYGTQTAIVVQSQDMAIRVQR"
2119..2571
/gene="csga"
2119..2571
/gene="csga"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-csga protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLIQVAPFAIVVSGSALAGSVPRQGGGGGGSGGSESTL
SIYSGVNNALALQSDPKRSPTTHQGFENGADVGSGSNTSIDLQNSPKNNATT
DQMGKSDITVSOTGHNALVNOTASDSSVLVHQVGFNNATNQR"

ORIGIN
Query Match      61.4%; Score 280; DB 1; Length 2889;
Best Local Similarity 77.4%; Pred. No. 1,66-69;
Matches 353; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

QY 1 ATGAAACTTTTAAAGTGGACGATTCGACGCAATCGTAGTTTCTGGCAGTCTTGCT 60
Db 2119 ATGAAACTTTTAAAGTGGACGATTCGACGCAATCGTAGTTTCTGGCAGTCTTGCT 2178
QY 61 GGCCTCCGTTCCACATGGGGGGGGCGCGGTATCATTAACGGCGCGCAATAGTTCGGC 120
Db 2179 GGTTCGTGTTCCGCAATGGGGGGGGCGCGG---TGGCGGGCGGGGAGCAGCTCCGCG 2235
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGGTGGCTGCTCTGCA 180
Db 2236 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGGTGGCTGCTCTGCA 2295
QY 2236 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGGTGGCTGCTCTGCA 2295
Db 181 AGCGATGCCGCTAAATATGATCAGCTGGTTACCGGTGTTGTTACCATGAATGGCAAT 240
QY 2296 AGCGATGCCGCTAAATATGATCAGCTGGTTACCGGTGTTGTTACCATGAATGGCAAT 2355
Db 241 GCGAGCCCGGGTGGCGATTAATGTAATCTTGAACGATCTCGAATGTTTCAGAATTAAT 300
QY 241 GCGAGCCCGGGTGGCGATTAATGTAATCTTGAACGATCTCGAATGTTTCAGAATTAAT 300
Db 2356 GTGGCCCGAGGCTCAGATTAACGACCATCATCTGACTCAAAACGGGCTTCAAAAACAC 2415
QY 301 GCCACATCGACGAGTGAACGCTAAAACTCCGATATTAAGTGGCCAAATACGGCGCT 360
Db 2416 GCCACATCGATCACTGAAACGGCAAAATTCGACATTTACTGTGAGCCAGTAATGTGGA 2475
QY 361 AATAACGGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Db 2476 CATTAACGGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 2535
QY 421 GGTTTTGGCAACAACGGCGCTTAACCAATTA 456
Db 2536 GGTTTTGGCAACAACGGCGCTTAACCAATTA 2571
```

```
RESULT 10
ECCSGABDG 4680 bp DNA linear ECT 07-JUL-2002
LOCUS
DEFINITION E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION X90754
VERSION X90754.1 GI:1147558
KEYWORDS csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
orfC gene.
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Hammar, M., Arngvist, A., Bian, Z., Olsen, A. and Normark, S.
Expression of two csg operons is required for production of
fibronectin- and congo red-binding curli polymers in Escherichia
coli K-12
Mol. Microbiol. 18 (4), 661-670 (1995)
JOURNAL 9614468
MEDLINE 8817489
PUBMED 2 (bases 1 to 4680)
REFERENCE
Hammar, M.
Direct Submision
Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
SWEDEN
FEATURES
source
Location/Qualifiers
1..4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/map="23.15 minutes"
1..3140
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="MC4100"
/db_xref="taxon:562"
3141..4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="N3110"
/db_xref="taxon:562"
<1..54680
/insertion_seq="IS2 (partial)"
complement(133..966)
/gene="csgG"
complement(133..966)
/gene="csgG"
/codon_start=1
/transl_table=11
/product="assembly /transport component in curli
production"
/protein_id="CA62277.1"
/db_xref="GI:1147559"
/db_xref="GOA:P52103"
/db_xref="SWISS-PROT:P52103"
/translation="MORFLVLAIVMLISGLTAPPEKAAAPTLMRAQSYKDLTHLPA
PTKGIKIVSVNIDDETQFQRPYPASNSFTAVPOSATMLVTALKDSRMPFLPREGIQ
NLNERTKIRAAQNGTVAINNRIPLQSLTAANIMVSGIIGVSNKSGGVGARVYG
IGADTQVLOLDIVNLRVNVSTGEILSVNTSKTILSYEVQGVPRPIQYORLLEGB
VGYTSNPVNLCLMSAIEFQVIFLINDIDRGLMDLONKAEKRONDILVTKRHNSVPE
S"
complement(993..1409)
/gene="csgF"
complement(993..1409)
/gene="csgF"
/codon_start=1
/transl_table=11
/product="assembly /transport component in curli
production"
/protein_id="CA62278.1"
/db_xref="GI:1147560"
```

```

/db_xref="SWISS-PROT:P52104"
/translation="MRVKAHVAVLLMLISPLSNAGMTFOFRNPNFGANNNGAFILNS
AOANSTKIDPSVNDPFEITPESALDNFTQAIQOQIIGLLSINTGKPSRMVNTIYIV
DIARIDQQLQNLVDRKTQSTIQVSGIQQNNSTDF"
complement(1434..1823)
/gene="csge"
complement(1434..1823)
/gene="csge"
complement(1434..1823)
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translation="MKRYLWIVAAEFLLAAGNLAAVEVEVPELLTDHTVSGIDHDFY
RPSDMSDDYGNLTINERPSARWGSMTITVNDVIFQPLFLKRDPEKTVVAFAL
IQTEALNRQNLQALLSTGDLADEF"
complement(1828..2478)
/gene="csge"
complement(1828..2478)
/gene="csge"
complement(1828..2478)
/notice="necessary for transcription of the csgeA operon"
/transl_table=1
/product="putative regulatory protein"
/protein_id="CA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translation="MENEVHSHGHITLLITKSSIQATYALQHKQSLATNGKLNHIO
RSLDISGSITLLDMEADKLIHYMDTISRKNNTKILLNTPEPYRDIENMP
HNGFYSEDEDERVNGILOGVLRGECEYOKLASYLTHSGRNTSBSALLTHEE
KEILNKLRIGANNEIARSIFISENTVTHLYNFKLIAVNRQVANSMDNINR"
complement(2626)
/notice="transcriptional start site for csgeDFG operon"
3140
/notice="transcriptional start site for csgeA operon"
3233..3688
/gene="csge"
3233..3688
/gene="csge"
3233..3688
/codon_start=1
/transl_table=1
/product="nucleation component of curlin monomers"
/protein_id="CA62281.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translation="MKKKLFPMLITLIGAPQIAAAGYDLANSEYNFAVNELSKSSFN
QAALIGAGTNNASQILROGSKLAAVVAQSSNRKIDTGDYINLAVIDQASANDA
SISQAYAGNTMIIQKSGNKANITQYCTOKTAIVQRCQMAIRVQR"
3729..4184
/gene="csge"
3729..4184
/gene="csge"
3729..4184
/notice="major subunit"
/codon_start=1
/transl_table=1
/product="curlin/CsgA protein"
/protein_id="CA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="MKLLKVAIAIAIVFSGSALAGVAVPOYGGGNGHGGGNNSGPNSF
LNIYQGGNSALALQTDANSDLTITQHGSGGADVGQSGDSSIDLTORGFNGSAT
LIDWNGKNSMTYKQFGGNGAIVDQTASNSVNTVQVFGNNAATAHQV"
3729..3788
/gene="csge"
3729..3788
/gene="csge"
4243..4575
/gene="orfC"
4243..4575
/gene="orfC"
sig_peptide
gene
CDS

```

```

/codon_start=1
/transl_table=1
/product="putative curli production protein"
/protein_id="CA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="MNTLLLAALSSQIFNTQGGDYVITPEVTLNOSLCRVOIL
SLRSGSGGQGTQKQKTLISLPANQPIALTKLSINISIPDRVKIVTVVSDGSLHSQQ
MPPSSKXS"

```

```

ORIGIN
Query Match 52.3%; Score 238.4; DB 1; Length 4680;
Best Local Similarity 70.2%; Pred. No. 1.6e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGTGGAGCATTCGACGCAATGTGTTCTGCGAGTCTGAGT 60
DB 3729 ATGAACTTTAAAGTGGAGCATTCGACGCAATGTGTTCTGCGAGTCTGAGT 60
QY 61 GGGCTGTTCCACAAATGGGCGGCGGCTGAATCAATACGGGCGGCAATGTTCCGCG 120
DB 3789 GGTGTTGTTCTCAGTACGCGCGCGGCTGAACGAGTGTGCGGTAAATAGCGCG 3848
QY 121 CGGACTCAACGTTGACATTTATCAATACGTTGCGTACGCTGCGTCTGCA 180
DB 3849 CCAAACTTCAAGCTGAACATTTACGATACGCTGCGGCTGAATCTGCACTGCTGCA 3908
QY 181 AGGATGCCGTAATATATATATGATGAGCTGTTACCGGTTGTTACCGCAATGCGACAT 240
DB 3909 ACTGATGCCGTAATATATATGATGAGCTGTTACCGGTTGTTACCGCAATGCGACAT 3968
QY 241 GCAAGCCAGGCGGCGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 3969 GTTGTGACGCGCTCAGATGACAGTCAATGATGACAGTCAATGATGACAGTCAATGAT 4028
QY 301 GCGACCATGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
DB 4029 GCTACTCTTGAATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4088
QY 361 AATAAGCGCGCGCTGTTAATGACACCGCATCTGATTCACGCTGAATGAGTCAAGTT 420
DB 4089 GCGAAGCGGCTGAGTGAATGACACCGCATCTGATTCACGCTGAATGAGTCAAGTT 4148
QY 421 GGTGTTGCAACGAGCGGCTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAG 456
DB 4149 GGTGTTGCAACGAGCGGCTGAATGAGTGAATGAGTGAATGAGTGAATGAGTGAATGAG 4184

RESULT 11
AE000205 10346 bp DNA linear BCT 01-DEC-2000
LOCUS Escherichia coli K12 MG1655 section 95 of 400 of the complete
DEFINITION genome.
ACCESSION AE000205 U00096
VERSION Escherichia coli K12 MG1655 section 95 of 400 of the complete
KEYWORDS genome.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
REFERENCE Escherichia coli K12
AUTHORS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 10346)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5311), 1453-1474 (1997)
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL MEDLINE
PUBMED 9278503
PUBMED 9742617
REFERENCE 2 (bases 1 to 10346)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,

```

University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 3 (bases 1 to 10346)
 Blattner, F.R.
 Direct Submission
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 4 (bases 1 to 10346)
 Plunkett, G. III.
 Direct Submission
 Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: mark@camber.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>): *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source
 1..10346
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 complement(1..48)
 /note="IS3"

repeat_region
 1..5
 /note="factor Sigma70; predicted +1 start at 1094689"

promoter
 5..32
 /note="factor Sigma70; predicted +1 start at 1094716"

gene
 69..392
 /gene="b1028"
 69..392
 /gene="b1028"
 /function="orf; Unknown"
 /note="0107; This 107 aa ORF is 27 pct identical (0 gaps)
 to 58 residues of an approx. 188 aa protein PAS2_CANAL
 SW: P43098"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC74113.1"
 /db_xref="GI:1787266"
 /translation="MEDGVQAMRDYLAGLDIASPEHOYLVNVTAKSEVAPSIKENIS
 LHLTHVAMTESLDITFLNMPYVAFLEISNRPYIGNMLNDPAGVDQQRVHMCRAFSFSD
 AKVPK"
 275..299
 /gene="b1028"

promoter

gene
 389..1375
 /note="factor Sigma32; predicted +1 start at 1094983"
 /gene="ycdU"
 /note="synonym: b1029"

CDS
 389..1375
 /gene="ycdU"
 /function="orf; Unknown"
 /note="0328; This 328 aa ORF is 19 pct identical (4 gaps)
 to 191 residues of an approx. 288 aa protein Y320_MYCE
 SW: P47562"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC74114.1"
 /db_xref="GI:1787267"
 /translation="MIPDYLTFRFDKRLIYIYAGILIGYWKNAQFTPSBDI
 GVSGIALVLYNFIIDLKXWMAKCTKNDISFWRKKXKHLEFLTOPVAGFIS
 LIMSASMGVGLVLPGLVPLVPLSLISLGPVIFLFPIMRTSYKQVALSVAKKVKY
 LITRYVLVSCISTVNVNLITSPISRNDSFTVTEGOMTFPSIIALLICGVVNLNF
 FLRSKRYAPLGRFLQRIIDLPFSESNALSTFPKPKMLALFILIVYEVMTITVSVL
 AILVEMKIMEAIFLCTVPCLIYFFYCFPLMHNDPMACDMTFWGHFNK"
 1253..1280
 /gene="ycdU"
 /note="central position to predicted promoter: -104"
 /bound_moiety="Deor predicted site"

protein_bind
 1253..1280
 /gene="ycdU"
 /note="No predicted promoter"
 /bound_moiety="Deor predicted site"

promoter
 1337..1365
 /gene="ycdU"
 /note="factor Sigma70; predicted +1 start at 1096049"

gene
 1494..1745
 /gene="b1030"

CDS
 1494..1745
 /gene="b1030"
 /function="orf; Unknown"
 /note="083; This 83 aa ORF is 25 pct identical (7 gaps)
 to847 residues of an approx. 616 aa protein ALBU_Rabbit SW:
 P49065"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC74115.1"
 /db_xref="GI:1787268"
 /translation="MGGTRCQVYTKKRPVLVYLFKYGEGGIDLSRPGCPAHAC
 VHGISNMLSPVVDPRSGLLIPVCAIHEKKARFVRALL"
 1513..1926
 /gene="ycdV"
 /note="synonym: b1031"

CDS
 1513..1926
 /gene="ycdV"
 /function="putative structure; Not classified"
 /note="0137; This 137 aa ORF is 27 pct identical (1 gap)
 to47 residues of an approx. 80 aa protein RS37_YEAST SW:
 P05759"

gene
 205759
 /codon_start=1
 /transl_table=1
 /product="putative ribosomal protein"
 /protein_id="AAC74116.1"
 /db_xref="GI:1787269"
 /translation="MCNIRKKSYPFRSSLSNMAVNGGLTRFAPRAGSPITATVCT
 GQRQSLPGRGFSPPRCVQTKKKPVLSYLFKYGEGGIDLSRPGCPAHACVHGL
 SNMLSPVVEPRSGLLIPVCAIYEKKARFVRALL"
 complement(2111..2198)
 /gene="serX"
 /note="synonym: b1032"
 complement(2111..2198)
 /gene="serX"
 /product="tRNA-Ser"
 /note="anticodon: GGA; CG Site No. 17878"
 /function="RNA; tRNA"

gene
 2269..2296

trna
 2269..2296

promoter

promoter /note="factor Sigma70; predicted +1 start at 1096980"
2338.2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355.2383
gene /note="factor Sigma70; predicted +1 start at 1097067"
2393.3370
/gene="ycdM"
/note="synonym: b1033"
2393.3370
CDS
/gene="ycdM"
/function="putative enzyme; Not classified"
/note="0325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPR_A_CORGL
SM: P45637"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="MRINFAORSGVSMIIFFYHPFDYQWTEALRKALIQAVRYAM
KSGDMSADYALVWHPVPEMLGRDLKAFALGAGVDLSLQAHPEMLNPVPLFR
LEDTGSGEOMQEVAVSYVLMFRFDYVITQNSHWPPLPEYHREDFTIGLGAVL
GSRVASLQTFMRPLRCWSRTKSPGVQSFAGRELSFLSQCEVLINLPETV
GIINQDLLEKLPDGAVALNLARGVHVEDDLADSGKVGAMLDVFNREDLPESP
LMQHPVITTPHVAATRPAAVEYISRTIAQLEKGRVCGVDARQY"
3278.3308
/gene="ycdM"
/note="factor Sigma70; predicted +1 start at 1097992"
3383.3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396.3423
promoter
promoter
promoter

Query Match 52.3%; Score 238.4; DB 1; Length 10346;
Best Local Similarity 70.2%; Pred. No. 1.7e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
Db 1 ATGAACTTTAAAGTGCAGCATTCGAGCATCTGAGTCTCTGCT 60
8993 ATGAACTTTAAAGTGCAGCATTCGAGCATCTGAGTCTCTGCT 60
Qy 61 GCGCTGTTCCACAATGGGGCGGCGGATATCATTAAGCGCGCAATGTTCCGCC 120
9053 GGTGTTGTTCTCAGTACGGCGCGGCGGATATCATTAAGCGCGCAATGTTCCGCC 120
Db 121 CCGGCTCAAGCTTATGAGATTTATCATGAGTTCGCTACGCTGCGCTGCTCTGCA 180
9113 CCAATTTGAGCGAATTTACAGTACGCTGCGGCTGCTGCACTTCTGCTGCA 180
Qy 181 AGCGATGCCGTAATATGATCAGCTGCTTACCGCTGTTGTTACCCATGAAATGGCAT 240
9173 ACTATGCCCTTACTGACTTATCTATTAACGATGCGCGGTAATGTTGTCAGAT 240
Db 241 GCAGCGCGGCTGCGATATGATATGATGATGATGATGATGATGATGATGATGAT 300
9233 GTTGTGCGGCTGCGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GCCACCATCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
9293 GCTACTCTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Db 361 AATAAGCCCGGCTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
9353 GGCACCGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 421 GGTTCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 480
9413 GGTTCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 480
Db 9413 GGTTCGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 480

RESULT 12
D90741 15047 bp DNA linear BCT 25-DEC-2002
LOCUS D90741
DEFINITION Escherichia coli K12 genomic DNA. (23,772 - 24,000 min).

ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
csgA; ycdB; cys; nov; mdcG.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
REFERENCE 1
AUTHORS Oshima, T., Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A.,
Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K.,
Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
Saito, Y., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y., Yano, M., and Horiuchi, T.
A 716-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)
8905232
JOURNAL MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL Unpublished
AUTHORS Mori, H.
3 (baes 1 to 15047)
REFERENCE
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (29-JUN-1996) Hirotsada Mori, NARA Institute of Science
and Technology, Res. & Edu. Center for Genetic Info., 8916-5
Takayama, Ikoma, Nara 630-01, Japan
(E-mail: hmori@gtc.aist-nara.ac.jp, Tel: 81-7437-2-5660,
Fax: 81-7437-2-5669)
COMMENT
Project:
The Japan E. coli genome DNA sequencing project
Group:
The Japan E. coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A.,
Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,
Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y., and Yano, M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotsada Mori
Address: NARA Institute of Science and Technology,
Ikoma, 630-01, Japan
E-mail: hmori@gtc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http://bsw3.aist-nara.ac.jp.
Location/Qualifiers
1.15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/note="Nucleotide position 1097101-1112147 from the

FEATURES

source

initiation site of ThrA (0 min.).-This clone is from Kohara lambda minisec library."

1..276
/note="unnamed protein product; ORF_ID:0230#5"
/codon_start=1
/transl_table=1
/protein_id="BAJ35822.1"
/db_xref="GI:4062605"
/translation="DLFPSSSENAIIFPAKPLMLRFLVLIEVMMITLVSVATLVE
WRIPFAVFLCYVCLLYFFCYCFMLWMDPMACDMYFRMGHFNK"
1152..2090
/note="ORF_ID:0231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=1
/product="Hypothetical 33.0 kd protein in prob-proA
intergenic region."
/protein_id="BAJ35823.1"
/db_xref="GI:4062606"
/translation="WDIIFYHPTPTOMWIALRKAIPOARVRAKSGDNDSDAIV
WHPVEMLAGDKAVFALGAGVDITLSKLAHPMLNPSPFLPRLBTGSGEQMRY
AVQVLMFRFRDDRIQNSHWPLEPHREDFITGILAGVLSKVAOSLQWRF
PLRCWSTRKSMGPVQSPAGRELSAFLSQCRVJLNPTEVGIINQDLKLPD
GAYLNLARGVHVEDDLAALDSGKGMALDVFNREPLPESPFLMQHPVITTPHV
AATRPAAVEYISRTIAQLKGRVCGQVDRAGY"
2145..2882
/note="unnamed protein product; ORF_ID:0231#2"
/codon_start=1
/transl_table=1
/protein_id="BAJ35824.1"
/db_xref="GI:4062607"
/translation="WYVDLHMHTVASTHAYSLTSDIYAOKKGIKLFATIDGPPM
EDAPHHMFRIMRIWPRVVDGVLIRGIEANIKVDSIICSGKPFSDLIINAGFHE
FVRPHDPAITQAMITATISGNVHIISHPGPKYIDVKAVAEAAAGQVALEPNS
SFLSRKSGEDNCREVAARVADAGVWALGSDSHATFTMGEBECLKILDVDPPEP
ILNVSPRLNLFLESGMAPIAEFADL"
2906..3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=1
/product="Hypothetical protein H11543"
/protein_id="BAJ35825.1"
/db_xref="GI:4062608"
/translation="WNEPSITCRVIGSYRQPODPLVPLFLIRECKLANPBLEQ
DELIRLOKSCDMQVSAVYALFTIGDCAFPPIRSANVEATEVAFISEKMP
ADTPADHIGILLAAWLEQSTEDSESLTLPSEYLLPWCGAFLGKVAHATTPTW
RTMAPLTRDAISAMWDELEBDEE"
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"
/codon_start=1
/transl_table=1
/protein_id="BAJ35826.1"
/db_xref="GI:4062609"
/translation="WCSISFMAAFSAIRKGNLILSLAITTGILSGIWVAVSLGL
SMAGFICTAYFACOGGKGLAISATLLSGVWAMWIIYIGSLAPLETILGYVITG
IVATLMCTQAKOLISFTYGTFIGACATPAQGDWKLVPSELALGLITGYMKNSGLM
LAARSATKARKEQETKNA"
complement (4117..4950)
/gene="cs8g"
/complement (4117..4950)
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs8g precursor."
/protein_id="BAJ35827.1"
/db_xref="GI:4062610"
/translation="MORLFLVAVWLISGLITAPPKEARPTLMPRAGSYDLTHLPA
PTGKIPASVNIODETGQFKRYPASNPATVPOSATVALLVALLDSRVPFLDEQGLQ
NLNBRKTIIRAAOENGVAINNRIPLQSLTAANIMVEGSIIGESNVSGFVGARVFG

IGADTOYOLDIOAVNLRVNVSTGEILSVSTKTLTISYEQAGVFRIDYORLLEGE
VGYTSNRPVWCLMSALETGYIFLINDIGRLMDLQNKARQNDILVKYRHSVPPE
S"
complement (4977..5393)
/gene="cs8g"
complement (4977..5393)
/gene="cs8g"
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs8g precursor."
/protein_id="BAJ35828.1"
/db_xref="GI:4062611"
/translation="WRVHAVALMLISPLSWAGTWTFQFPNPNFGNPNNGAFLLNS
AQONSYKDPBYNDPFIETPSALDNFQALQSOIILGGLLSINITGKRGKRVYTDYIV
DIARNDQQLQNLVDRKTGTSITQVSGLIQNNSTDF"
complement (5418..5807)
/gene="cs8g"
complement (5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs8g precursor."
/protein_id="BAJ35829.1"
/db_xref="GI:4062612"
/translation="MKRYLRWIVAAEPLFAAGNLHAVEVYRGLTDTHTVSSIGDPY
RASPDKESDYTGTLTINERPSAWGSMWITTVNQDVIPOFPLPLKRDPEKTVVFL
IQTEBALNRKQINQALSTGDLADEP"
complement (5812..6462)
/gene="cs8g"
complement (5812..6462)
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=1
/product="Probable cs8g operon transcriptional regulatory
protein."
/protein_id="BAJ35830.1"
/db_xref="GI:4062613"
/translation="WNEVBSIHGHTLLITKSLQATPALLQHLKQSLATIGKLMNQ
RSLDDISSGIIILDMNEADKLIHYQDTLSRKNNIKILMLTPEDYPYRDIENMP
HINGVFYSMEQERVVNGLOGLVEGECYFOKLASVYLTHSGNRYNSTESALTHRS
KEILNKLRIQIGSNNEIARSLFISENTVKTLYNLFKLIAVKNRTQAVSAMNDILRR"
7217..7672
/gene="cs8g"
complement (7217..7672)
Query Match 52.3%; Score 238.4; DB 1; Length 15047;
Best Local Similarity 70.2%; Pred. No. 1.7e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGATTGCGACAAATCGTAGTTCTGCGAGTCTTGCT 60
DB 7713 ATGAACTTTTAAAGTGGCAGATTGCGACAAATCGTAGTTCTGCGAGTCTTGCT 7712
QY 61 GCGCTGCTTCCACAAATGCGCGCGCGCGGCGTAATACATACGCGCGCGCAATAGTCCGCG 120
DB 7773 GGTGTGTCTCTCGTACGCGCGCGCGCGTAACACGCGTGTGCGGTAATATATGACGCG 7832
QY 121 CCGGACTCAAGCTTGACATTTATCAGTACCGGTTCCGCTAACGCTGCGCTTCTCGCA 180
DB 7833 CCAATTTCTAGCTGACCATTTACCATGACGCTGCGGTAATCTGTGACCTTCTCGCA 7892
QY 181 AGCAGTCCCGTAATATGATCAGCTGCTTACCCGTGTTGTTACCAATGAATGCAAT 240
DB 7893 ACTGATGCCGTAATCTGACTTACTATTAACACGAGATGCGCGGTAATGATGTCAGAT 7952

QY 241 GCAGCGCAGCGGTCGGATATAGTACTATGACTGACTGAAATGTTTCAGAAATAT 300
 DB 7953 GTTGTTCAGGCTCAGATGACAGCTCATCTGACCCAACTGGCTTCGTAAACAGC 8012
 QY 301 GCCACCATGACGACTGACGCTAAACCTCCGATTTACTGTGCGCCAAATCGGGGT 360
 DB 8013 GCTACTCTTGATCAGTGAACGGGCAAAATCTGATGACGGTTAAACAGTTGGGTGT 8072
 QY 361 AATACGCCGCGCTGCTGATATACAGCCGATCTGATTCAGCGTAAATGTCGTACGTT 420
 DB 8073 GGCACAGGCTGTCAGATGACGATGACATCTAACTCTCCGTCACAGTCACTGAGTT 8132
 QY 421 GGTTTGGCAACAACGCCACGCTTACACGATTTAA 456
 DB 8133 GGTCTTGGTACAAACGCCACGCTCATCAGTACTTA 8168

RESULT 13

CFS15701 2920 bp DNA linear BCT 24-JUN-2003
 LOCUS Citrobacter freundii csb gene, csbA gene and csbD gene.
 DEFINITION A515701
 ACCESSION A515701.1 GI:31790495
 VERSION csbA gene; csbD gene; curlin-csbA protein; nucleation
 KEYWORDS component of curlin monomers; regulatory protein.
 SOURCE Citrobacter freundii
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Citrobacter.
 REFERENCE 1 Zogaj, X., Bokranz, W., Nimtz, M. and Romling, U.
 AUTHORS Production of Cellulose and Curli Fimbriae by Members of the Family
 TITLE Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
 JOURNAL Infect. Immun. 72 (7), 4151-4158 (2003)
 AUTHORS 2 (bases 1 to 2920)
 TITLE Romling, U.
 JOURNAL Direct Submission
 Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
 Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES
 source Location/Qualifiers
 1..2920
 /organism="Citrobacter freundii"
 /mol_type="genomic DNA"
 /isolate="Pec4"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:546"
 /country="Germany"
 /gene="csbD"
 complement(211..861)
 /gene="csbD"
 complement(211..861)
 /function="necessary for transcription of the csbA
 operon"
 /codon_start=1
 /transl_table=11
 /product="putative regulatory protein"
 /protein_id="CAD56673.1"
 /db_xref="GI:31790496"
 /translation="MREKVSHTGHTLLITKPSLQATALLQKLSALNGLKLNQ
 RSLDSSSCIVLDMKADKLIHWQDLNRKNNIKTLINLPDYPDINMP
 HINCFYATDEORVVGLOGVGEYSOKLASYLITSGNRYNSTESALLTHRE
 KEIINKIRIGASNIETARSLEISNTVKTHTLVLFKIAVKORTQAVSWANNLR"
 1624..2082
 /gene="csbB"
 1624..2082
 /gene="csbB"
 /function="nucleator"
 /codon_start=1
 /transl_table=11
 /product="nucleation component of curlin monomers"
 /protein_id="CAD56674.1"
 /db_xref="GI:31790497"

ORIGIN

Query Match 52.0%; Score 237.2; DB 1; Length 2920;
 Best Local Similarity 71.7%; Pred. No. 3, 5e-57;
 Matches 327; Conservative 0; Mismatches 123; Indels 6; Gaps 1;

/translation="MKKKLIFMFMFTMLGVGIAAATTSYDLANSEYNFAVNELEKSSF
 NQAAIQDQVGNANRTOGGSKLSVTSOEGSGNRAKTQDTSYNPAVYIDQSGSSND
 ASIKQSGYGNPAVYIIOKSGSNKANITQYGTOKTAAVVVQRSQAAIVTQR"
 2123..2572
 /gene="csbA"
 2123..2572
 /gene="csbA"
 /function="curli subunit"
 /codon_start=1
 /transl_table=11
 /product="curlin-csbA protein"
 /protein_id="CAD56675.1"
 /db_xref="GI:31790498"
 /translation="MKLIKVAAPAAIVVSGSLAGVVPQGNHGGGSNYPDSSL
 IYQGSNNSNMAIQSDARKSDVITTHGNGAIVVGGQADDSITSLKQTFONSATID
 QMAKRAIDSVTFGRNPAIVNQTASDSNVLIQVGFNNATANQH"

QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATGTAAGTTCTGCGAGTCTGCT 60
 DB 2123 ATGAACCTTTAAAGTGGCAGCATTCGACGAATGTAAGTTCTGCGAGTCTGCT 2182
 QY 61 GCGGTGCTTCAAAATGGGCGCGCGGTATCATACCGCGCGGCAATAGTCCGCG 120
 DB 2183 GGTGTGTTCCGCAATGGGCGCGC-----ATCATCATGATGGTGGGCAATTAATGCG 2236
 QY 121 CCGGATCTCAAGTTGACATTTATAGTACGGTTCGCTTACCGTCCCTGCTCGCA 180
 DB 2237 CCAAGTCTTCTCACTAGATATCAACATATGGGTCAACCACTCGCAAAATGCTCTGCA 2236
 QY 181 AGCGATGCCCTTAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 DB 2297 AGTGAATGCCCTTAATATGATGATGATGATGATGATGATGATGATGATGATGAT 2356
 QY 241 GCAAGCCAGGGTGGGATATAGTATGATGATGATGATGATGATGATGATGATGAT 300
 DB 2357 GTTGGCCAGGGTGGGATATAGTATGATGATGATGATGATGATGATGATGATGAT 2416
 QY 301 GCCACCATGACGACTGACGCTAAACCTCCGATTTACTGTGCGCCAAATCGCGGT 360
 DB 2417 GCCACTATCGATCAGTGAATGACAAATGCTGATATGACGCTGACCGCTTCGAGAACGT 2476
 QY 361 AATACGCCGCGCTGCTGATATACAGCCGATCTGATTCAGCGTAAATGTCGTACGTT 420
 DB 2477 GGCACAGGCTGTCAGATGACGATGACATCTAACTCTCCGTCACAGTCACTGAGTT 2536
 QY 421 GGTTTGGCAACAACGCCACGCTTACACGATTTAA 456
 DB 2537 GGTCTTGGTACAAACGCCACGCTCATCAGTACTTA 2572

RESULT 14

A814811 456 bp DNA linear PAT 05-DEC-2003
 LOCUS Sequence 15 from Patent WO03064446.
 DEFINITION A814811
 ACCESSION A814811
 VERSION A814811.1 GI:39104001
 KEYWORDS Escherichia coli
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE 1 Bioerck, L., Olsen, A., Wikstroem, M. and Herwald, H.
 AUTHORS Peptides
 TITLE Patent: WO 03064446-A 15 07-AUG-2003;
 JOURNAL Hansa Medical Research Aktiebolag (SE)
 FEATURES
 source Location/Qualifiers
 1..456

CDS

```

/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..456
/note="unnamed protein product"
/codon_start=1
/translation_table=11
/protein_id="CAE83541.1"
/db_xref="GI:39104002"
/translation="MKLIKVAIAIVSGSALAGVIVQYGGGNGHGGGNSGPNSE
LNIYQYGGNSALALQTDARNSDLITTHGGGNGADVQGSDDSSIDLTRGFGNSAT
LDQWNGKNSMTVAKQFGGNGAAVDQTAHSNSVNVTVQFGNNATAHOY"

```

ORIGIN

```

Query Match      51.9%; Score 236.8; DB 6; Length 456;
Best Local Similarity 70.0%; Pred. No. 4.1e-57;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGCGAGTGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGAAGCAATTCGACGATCGTAGTTCTCGGTAAGCGCTTGCA 60
QY 61 GCGCTGTTCCAAATGGGCGCGCGGTAATCATAAACGGCGGCAATAGTTCCGCGC 120
DB 61 GGTGTGTTCCCTCAGTACGCGCGCGGTAACACAGGTGGTGGCGGTAAATAGCGGC 120
QY 121 CCGGATCGACGTTAGATTATTCAGTACGTTCCGCTTAAGCTGCGCTGCTCTGCA 180
DB 121 CCAATTTGAGCTAAACATTTACAGTACGTTGCGGTAACCTGCACTTGCTCTGCA 180
QY 181 AGCGATGCCGCTAAATATGATCAGCTGGTACCCGTTGTTGTTACCATGAATGACAT 240
DB 181 ACTGATGCCGCTAAATATGATCAGCTGGTACCCGCTTAACCATGAATGACAT 240
QY 241 GCAAGCCAGGTCGCGATTAATGATTAATGAACTGACTCAGAAATGTTTCAGAAATAT 300
DB 241 GTTGTGTCAGGCTCAGTACGAGCTCAATCGATCGACCAACGTCGCTGCTTAACGC 300
QY 301 GCCACCATGACGATGGAACGCTAAATCCGCTATTAATCTGTGGCCATACGCGCGT 360
DB 301 GCTACTCTTGATCACTGGAACGCGCAAAATTTGAAATGACGCTTAAACAGTTCGTGCT 360
QY 361 AATAAGCCGCGCTGTTAATACAGCCGATCTGATTCAGGTAATGATGCTGCTGAGTT 420
DB 361 GCGAAGCGTGTGCTGAGTGAACGACGATCTGATCTCTCCGTAAGTACTGAGTT 420
QY 421 GGTTTGGCAACAACGCGCAGCGCTAACCAATTA 456
DB 421 GCGTTTGGTAAACAACGCGCAGCGCTCATGACTTA 456

```

RESULT 15

LOCUS ECOCSSGA 648 bp DNA linear BCT 13-JUL-1993

DEFINITION Escherichia coli curlin subunit (csgA) gene, complete cds.

ACCESSION L04979

VERSION L04979.1 GI:290424

KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.

SOURCE Escherichia coli

ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Escherichia; Escherichia.

REFERENCE 1 (bases 1 to 648)

Olseu, A., Arny, J., A., Hammar, M., Sukhopol, S. and Normark, S.

The RpoD sigma factor relieves H-NS-mediated transcriptional

repression of csgA, the subunit gene of fibronectin-binding curli

in Escherichia coli

Mol. Microbiol. 7 (4), 523-536 (1993)

JOURNAL MEDLINE 93311294

COMMENT 8459772

On Jun 11, 1993 this sequence version replaced gi:145630.

Original source text: Escherichia coli (sub_strain W3110, strain

FEATURES

source

```

Location/Qualifiers
1..648
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/feature_1fb="Kohara"
83..538
/gene="csgA"
83..538
/gene="csgA"

```

gene

```

/note="precursor protein"
/codon_start=1
/evidence=experimental
/translation_table=11
/product="curlin subunit"
/protein_id="AA23616.1"
/db_xref="GI:290425"
/translation="MKLIKVAIAIVSGSALAGVIVQYGGGNGHGGGNSGPNSE
LNIYQYGGNSALALQTDARNSDLITTHGGGNGADVQGSDDSSIDLTRGFGNSAT
LDQWNGKNSMTVAKQFGGNGAAVDQTAHSNSVNVTVQFGNNATAHOY"

```

CDS

```

/gene="csgA"
/evidence=experimental
143..535
/gene="csgA"
/product="curlin subunit"
/evidence=experimental

```

sig_peptide

```

143..535
/gene="csgA"
/product="curlin subunit"
/evidence=experimental

```

mat_peptide

```

143..535
/gene="csgA"
/product="curlin subunit"
/evidence=experimental

```

ORIGIN

```

Query Match      51.9%; Score 236.8; DB 1; Length 648;
Best Local Similarity 70.0%; Pred. No. 4.1e-57;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGCGAGTGTCTGCT 60
DB 83 ATGAACTTTTAAAGTGAAGCAATTCGACGATCGTAGTTCTCGGTAAGCGCTTGCA 142
QY 61 GCGCTGTTCCAAATGGGCGCGCGGTAATCATAAACGGCGGCAATAGTTCCGCGC 120
DB 143 GGTGTGTTCCCTCAGTACGCGCGCGGTAACACAGGTGGTGGCGGTAAATAGCGGC 202
QY 121 CCGGATCGACGTTAGATTATTCAGTACGTTCCGCTTAAGCTGCGCTGCTCTGCA 180
DB 121 CCAATTTGAGCTAAACATTTACAGTACGTTGCGGTAACCTGCACTTGCTCTGCA 180
QY 181 AGCGATGCCGCTAAATATGATCAGCTGGTACCCGTTGTTGTTACCATGAATGACAT 240
DB 203 CCAATTTGAGCTAAACATTTACAGTACGTTGCGGTAACCTGCACTTGCTCTGCA 262
QY 241 GCGATGCCGCTAAATATGATCAGCTGGTACCCGTTGTTGTTACCATGAATGACAT 300
DB 263 ACTGATGCCGCTAAATATGATCAGCTGGTACCCGTTGTTGTTACCATGAATGACAT 322
QY 301 GCGATGCCGCTAAATATGATCAGCTGGTACCCGTTGTTGTTACCATGAATGACAT 360
DB 383 GGTGTGTTCCCTCAGTACGCGCGCGGTAACACAGGTGGTGGCGGTAAATAGCGGC 442
QY 361 AATAAGCCGCGCTGTTAATACAGCCGATCTGATTCAGGTAATGATGCTGCTGAGTT 420
DB 443 GCGAAGCGTGTGCTGAGTGAACGACGATCTGATCTCTCCGTAAGTACTGAGTT 502
QY 421 GGTTTGGCAACAACGCGCAGCGCTAACCAATTA 456
DB 503 GCGTTTGGTAAACAACGCGCAGCGCTCATGACTTA 538

```

Search completed: March 15, 2004, 22:50:17

Job time : 1965.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456
Sequence: 1 atgaacttcaaaagtcgc.....ccacggcctaccagratraa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 337363 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: geneseqn19808:*
- 2: geneseqn19908:*
- 3: geneseqn20008:*
- 4: geneseqn20018:*
- 5: geneseqn20028:*
- 6: geneseqn20038:*
- 7: geneseqn20048:*
- 8: geneseqn20058:*
- 9: geneseqn20068:*
- 10: geneseqn20078:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	3	AAC64626
2	412.8	90.5	456	2	AAQ87467
3	412.8	90.5	456	2	AAQ87467
4	412.8	90.5	456	2	AAQ87467
5	385.6	84.6	456	3	AAC64629
6	363.2	79.6	456	3	AAC64629
7	361.6	79.3	456	3	AAC64629
8	361.6	79.3	456	3	AAC64629
9	361.6	79.3	456	3	AAC64629
10	360	78.9	456	3	AAC64629
11	360	78.9	456	3	AAC64629
12	358.4	77.9	456	3	AAC64629
13	358.4	77.9	456	3	AAC64629
14	288.6	63.3	456	2	AAQ73066
15	288.6	63.3	456	2	AAQ73066
16	236.8	51.9	456	2	AAQ73066
17	236.8	51.9	456	2	AAQ73066
18	224.4	49.2	456	2	AAQ73066
19	162.4	35.6	369	2	AAQ73066
20	51.2	11.2	100	7	AAQ73066
21	50	11.0	78	3	AAQ73066
22	50	11.0	78	3	AAQ73066
23	48.2	10.6	78	9	AAQ73066

24	48	10.5	48	3	AAC64621
25	45	9.9	78	3	AAC64605
26	44.8	9.8	72	3	ACP36148
27	43.4	9.5	78	3	AAC64605
28	43.2	9.3	100	7	AAC64616
29	42.4	9.3	100	7	ACD68808
30	42	9.2	2000	7	ADA71938
31	35	7.7	349980	6	ABO81847
32	34	7.5	1344	2	AAQ09007
33	34	7.5	1344	3	AAQ09007
34	34	7.5	1344	3	AAQ09007
35	34	7.5	1344	3	AAQ09007
36	34	7.5	1344	6	AAQ29125
37	34	7.5	1344	6	AAQ29125
38	34	7.5	1344	6	AAQ29125
39	33	7.2	110000	6	ABA90521.17
40	32.8	7.2	3562	4	ABL06749
41	32.8	7.2	24081	7	AAQ54223
42	32.8	7.2	53101	7	AAQ54217
43	32.6	7.1	100	7	ACD68809
44	32.6	7.1	2553	3	AAA09500
45	32.4	7.1	901	4	ABL09481

ALIGNMENTS

RESULT 1	AAC64626	AAC64626 standard; DNA; 456 BP.
ID	AAC64626	
AC	AAC64626	
XX		
DT	26-FEB-2001	(first entry)
DE	AGFA:PT3#5 DNA sequence SEQ ID NO:19.	
XX		
KM	Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;	
KM	vaccine; immune response; immunogen; ds.	
OS	Salmonella enteritidis.	
OS	Escherichia coli.	
XX	Synthetic.	
PM	WO200060102-A2.	
XX		
PD	12-OCT-2000.	
XX		
PF	05-APR-2000; 2000WO-CA000356.	
XX		
PR	05-APR-1999; 99US-0127888P.	
XX		
PA	(UUYI-) UNIV VICTORIA.	
XX		
PI	White AP, Doran JL, Collison SK, Kay WW;	
XX		
DR	WPI: 2000-672631/65.	
XX		
PT	P-PSDB; AAB36350.	
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence	
PT	which encodes foreign epitope or antigen, expresses recombinant Agfa	
PT	protein useful for eliciting immune response in animal.	
XX		
PS	Disclosure; Page 137; 139pp; English.	
XX		
CC	The present invention describes a recombinant agfa gene (1) where a	
CC	segment of the gene has been replaced by a segment of a foreign DNA	
CC	sequence which encodes a foreign epitope or antigen. Also described are:	
CC	(1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended	
CC	assembly system of strains of Salmonella, Escherichia coli and	
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant	
CC	Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)	
CC	directing recombination of a recombinant gene into the chromosome of the	

homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrin subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrils are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;

Best Local Similarity 100.0%; Pred. No. 2.2e-139;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGCAGTGTCTGCT 60
DB 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGCAGTGTCTGCT 60
QY 61 GGGGTGTTTCCAAATGGGGCGCGCGGTATCATACGGCGCGCAATAGTCCGGC 120
DB 61 GGGGTGTTTCCAAATGGGGCGCGCGGTATCATACGGCGCGCAATAGTCCGGC 120
QY 121 CCGGACTCAAGCTTGGAGCATTTATCAGTGGTCCGCTACAGCGCGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGGAGCATTTATCAGTGGTCCGCTACAGCGCGCTGCTGCA 180
QY 181 AGCGATGCCCGTAATATGATCAGTGGTACCGGTGTTTACCAATGAGCAT 240
DB 181 AGCGATGCCCGTAATATGATCAGTGGTACCGGTGTTTACCAATGAGCAT 240
QY 241 GCAGGCCAGGGTGGGATTAATGATCTTGAATGAGTCAAGTGGTTCAGAAATAT 300
DB 241 GCAGGCCAGGGTGGGATTAATGATCTTGAATGAGTGGTTCAGAAATAT 300
QY 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATTTACTGCGCAATACGGCGGT 360
DB 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATTTACTGCGCAATACGGCGGT 360
QY 361 AATAACGCGCGGTGTTATCATACGCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAACGCGCGGTGTTATCATACGCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTITGGCAACACGCGCAGGCTAACAGATTTAA 456
DB 421 GGTITGGCAACACGCGCAGGCTAACAGATTTAA 456

```

RESULT 2

AA087467
ID AA087467 standard; DNA; 456 BP.

AC AA087467;

XX 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)

XX Agfa sequence.

XX *Salmonella*; Agfa; vaccine; genetic immunization; ds.

XX *Salmonella*.

Key Location/Qualifiers
FT CDS 1..454
FT /tag=a
FT /note="Agfa"

W09425598-A2.

10-NOV-1994.

26-APR-1994; 94NO-IB000207.

26-APR-1993; 93US-00054452.

(UVI-) UNIV VICTORIA INNOVATION & DEV CORP.
(KING/) KING J.

Key MW, Collinson SK, Clouthier SC, Doran JL;

WPI; 1994-358275/44.

P-PSDB; AAR74625.

Eliciting an immune response to *Salmonella* - using attenuated *Salmonella* strains, vector constructs, or compans. contg. fibrin type proteins.

Disclosure; Fig 7B; 95pp; English.

The DNA encodes the *Salmonella* Agfa protein. The DNA and isolated proteins are used in genetic immunization and vaccine compositions, respectively, to elicit an immune response to *Salmonella* in animals (e.g. food producing animals) and humans. (Updated on 25-MAR-2003 to correct FN field.)

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 2; Length 456;

Best Local Similarity 94.1%; Pred. No. 3.6e-125;

Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGCAGTGTCTGCT 60
DB 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGCAGTGTCTGCT 60
QY 61 GGGGTGTTTCCAAATGGGGCGCGCGGTATCATACGGCGCGCAATAGTCCGGC 120
DB 61 GGGGTGTTTCCAAATGGGGCGCGCGGTATCATACGGCGCGCAATAGTCCGGC 120
QY 121 CCGGACTCAAGCTTGGAGCATTTATCAGTGGTCCGCTACAGCGCGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGGAGCATTTATCAGTGGTCCGCTACAGCGCGCTGCTGCA 180
QY 181 AGCGATGCCCGTAATATGATCAGTGGTACCGGTGTTTACCAATGAGCAT 240
DB 181 AGCGATGCCCGTAATATGATCAGTGGTACCGGTGTTTACCAATGAGCAT 240
QY 241 GCAGGCCAGGGTGGGATTAATGATCTTGAATGAGTCAAGTGGTTCAGAAATAT 300
DB 241 GCAGGCCAGGGTGGGATTAATGATCTTGAATGAGTGGTTCAGAAATAT 300
QY 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATTTACTGCGCAATACGGCGGT 360
DB 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATTTACTGCGCAATACGGCGGT 360
QY 361 AATAACGCGCGGTGTTATCATACGCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAACGCGCGGTGTTATCATACGCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTITGGCAACACGCGCAGGCTAACAGATTTAA 456
DB 421 GGTITGGCAACACGCGCAGGCTAACAGATTTAA 456

```

RESULT 3

AAT74142

```

ID  AAT74142 standard; DNA; 456 BP.
XX  AC
XX  AAT74142;
XX  DT 25-MAR-2003 (revised)
XX  DT 29-SEP-1997 (first entry)
XX  DE Salmomella enteritidis 27655-3b agfa gene.
XX  KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX  OS Salmomella enteritidis.
XX  FH Key
XX  FT CDS
XX  FT Location/Qualifiers
XX  FT 1..456
XX  FT /tag= a
XX  FT /label= agfa_gene_fragment
XX  FT /transl_except= (pos:367..369,aa:Pro)
XX  PN US5635617-A.
XX  PD 03-JUN-1997.
XX  PF 26-APR-1994; 94US-00233788.
XX  PR 26-APR-1993; 93US-00054452.
XX  PA (UVI- ) UNIV VICTORIA INNOVATION & DEV CORP.
XX  PI Collinson SK, Kay WW, Doran JL;
XX  DR WPI:1997-309886/28.
XX  DR P-PSDB; AAW23570.
XX  PT Isolated Salmomella gene agfa - used for diagnosis of Salmomella or
XX  PT enteropathogenic bacteria of the Enterobacteria family.
XX  PS Claim 1; Col 19-112; 85pp; English.
XX  CC The present sequence represents an isolated agfa gene derived from
XX  CC Salmomella enteritidis 27655-3b. The nucleic acid can be used to provide
XX  CC diagnostic assays for Salmomella and/or enteropathogenic bacteria of the
XX  CC family Enterobacteria. It can also be used to provide proteins and
XX  CC antibodies which can be used for assays. The nucleic acid sequence can be
XX  CC used to provide probes or primers which can specifically hybridise to
XX  CC nucleic acid molecules from greater than 99% of Salmomella strains that
XX  CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
XX  CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
XX  CC correct PF field.)
XX  SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 2; Length 456;
Best Local Similarity 94.1%; Pred. No. 3.6e-125;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCGCATTCGACCAATCGTAGTTTCTGGAGTCTTGCTGCT 60
DB 1 ATGAAACTTTTAAAGTGGCGCATTCGACCAATCGTAGTTTCTGGAGTCTTGCTGCT 60
QY 61 GGCCTGTTCCACATGAGGCGGCGCGGCGGTATCATACGCGGCGGCATATGTTCCGGC 120
DB 61 GGCCTGTTCCACATGAGGCGGCGGCGGCGGTATCATACGCGGCGGCATATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTACGCTGCGCTTCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTACGCTGCGCTTCTCTGCA 180
QY 181 AGCGATCCCGCTAAATATGATCAGTACGCTTCCGCTGTTTCCATCCATGAATGCGACAT 240
DB 181 AGCGATCCCGCTAAATATGATCAGTACGCTTCCGCTGTTTCCATCCATGAATGCGACAT 240
QY 241 GCAGCGCAGGCTGCGATATATGATGATGAACTGACTGCAATGTTTCAAAATATAT 300

```

```

DB 241 GTAGCCAGGCTGCGATATATGATGATGAACTGACTGCAATGTTTCAAAATATAT 300
QY 301 GCCACATCGACGATGGAACGCTAAAACTCCGATATTAATCTGCGGCAATACGCGCT 360
DB 301 GCCACATCGACGATGGAACGCTAAAACTCCGATATTAATCTGCGGCAATACGCGCT 360
QY 361 AATTAACCGCGCTGCTGTTTATCAGACCGCATCTGATTCGACGCTAAATGTCGTCAGTT 420
DB 361 AATTAACCGCGCTGCTGTTTATCAGACCGCATCTGATTCGACGCTAAATGTCGTCAGTT 420
QY 421 GGTTTGGCAACACGCGCAGGCTAACGATTTAA 456
DB 421 GGTTTGGCAACACGCGCAGGCTAACGATTTAA 456

RESULT 4
AAC64617
ID AAC64617 standard; DNA; 456 BP.
XX  AC
XX  AAC64617;
XX  DT 26-FEB-2001 (first entry)
XX  DE Salmomella enteritidis Agfa DNA sequence SEQ ID NO:1.
XX  KM Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;
XX  KM vaccine; immune response; immunogen; ds.
XX  OS Salmomella enteritidis.
XX  PN WO200060102-A2.
XX  PD 12-OCT-2000.
XX  PF 05-APR-2000; 2000WO-CA000356.
XX  PR 05-APR-1999; 99US-0127888P.
XX  PA (UVI- ) UNIV VICTORIA.
XX  PI White AP, Doran JL, Collinson SK, Kay WW;
XX  DR WPI: 2000-672631/65.
XX  DR P-PSDB; AAB36341.
XX  PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX  PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX  PT protein useful for eliciting immune response in animal.
XX  PS Disclosure; Page 134; 139pp; English.
XX  CC The present invention describes a recombinant agfa gene (I) where a
XX  CC segment of the gene has been replaced by a segment of a foreign DNA
XX  CC sequence which encodes a foreign epitope or antigen. Also described are:
XX  CC (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
XX  CC assembly system of strains of Salmomella, Escherichia coli and
XX  CC Enterobacteriaceae for the production of fimbriae comprising recombinant
XX  CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX  CC directing recombination of a recombinant gene into the chromosome of the
XX  CC homologous species; (3) directing recombination of a recombinant gene
XX  CC back into the chromosome of the homologous species, replacing the native
XX  CC copy of that gene; and (4) eliciting an immune response in an animal,
XX  CC comprising separating an amino acid polymer comprising a recombinant Agfa
XX  CC protein containing a replacement segment or segments of foreign amino
XX  CC acid sequence or sequences grown on a Salmomella, B. coli or
XX  CC Enterobacteriaceae host cell, from the host cell and introducing the
XX  CC polymer into the animal in conjunction with a carrier or diluent. (I) is
XX  CC useful for the expression of recombinant Agfa protein which is useful for
XX  CC eliciting an immune response in an animal. In a fimbrial presentation
XX  CC system the heterologous antigens are presented in high numbers (up to
XX  CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX  CC immunogenicity and adhesion properties relevant for an efficient live

```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 3; Length 456;
 Best Local Similarity 94.1%; Pred. No. 3.6e-125;
 Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGACGATTCGACGATCGTAGTTCTCGAGAGCTCTGCT 60
DB 1 ATGAACCTTTAAAGTGGACGATTCGACGATCGTAGTTCTCGAGAGCTCTGCT 60
QY 61 GCGGTGTTCCAAATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGCG 120
DB 61 GCGGTGTTCCAAATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTTCTGCGAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTTCTGCGAA 180
QY 181 AGCGATGCGCCGTAATATGATCAGCTGTTACCCGTGTTTACCATGAAATGCGACAT 240
DB 181 AGCGATGCGCCGTAATATGAAACGACCATTAACGAGCGGTATGTAACGCGCCGAT 240
QY 241 GAGGCGCAGGGTGGGATTAATAGTATGAACTGATCGAGTCCGATTCGAAATAT 300
DB 241 GTAGGCGCAGGGTGGGATTAATAGTATGAACTGATCGAGTCCGATTCGAAATAT 300
QY 301 GCCACCATTCGACGATGGAAGCTTAAAGTCCGATATTAAGTCCGCAATACGGCGGT 360
DB 301 GCCACCATTCGACGATGGAAGCTTAAAGTCCGATATTAAGTCCGCAATACGGCGGT 360
QY 361 AATAAGCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTAATGTCGTCAGATT 420
DB 361 AATAAGCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTAATGTCGTCAGATT 420
QY 421 GGTGTTGGCAACAGCGCGCTAACCGATATTA 456
DB 421 GGTGTTGGCAACAGCGCGCTAACCGATATTA 456

```

RESULT 5
 AAC64629
 ID AAC64629 standard; DNA; 456 BP.
 XX
 AC AAC64629;
 XX
 DT 26-FEB-2001 (first entry)
 XX

DE Agfa: PT3#8 DNA sequence SEQ ID NO:25.

XX Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen; ds.
 XX

OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 OS

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WM;

XX
 DR WPI; 2000-672631/65.
 DR P-PSDB; AAB36353.
 XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 84.6%; Score 385.6; DB 3; Length 456;
 Best Local Similarity 90.4%; Pred. No. 3.1e-116;
 Matches 412; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGACGATTCGACGATCGTAGTTCTCGAGAGCTCTGCT 60
DB 1 ATGAACCTTTAAAGTGGACGATTCGACGATCGTAGTTCTCGAGAGCTCTGCT 60
QY 61 GCGGTGTTCCAAATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGCG 120
DB 61 GCGGTGTTCCAAATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTTCTGCGAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTTCTGCGAA 180
QY 181 AGCGATGCGCCGTAATATGATCAGCTGTTACCCGTGTTTACCATGAAATGCGACAT 240
DB 181 CTEGTTACCGGTGTTTACCATGAAATGCGACATGAGGTATGTAACGGCGCGAT 240
QY 241 GAGGCGCAGGGTGGGATTAATAGTATGAACTGATCGAGTCCGATTCGAAATAT 300
DB 241 GTAGGCGCAGGGTGGGATTAATAGTATGAACTGATCGAGTCCGATTCGAAATAT 300
QY 301 GCCACCATTCGACGATGGAAGCTTAAAGTCCGATATTAAGTCCGCAATACGGCGGT 360
DB 301 GCCACCATTCGACGATGGAAGCTTAAAGTCCGATATTAAGTCCGCAATACGGCGGT 360
QY 361 AATAAGCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTAATGTCGTCAGATT 420
DB 361 AATAAGCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTAATGTCGTCAGATT 420
QY 421 GGTGTTGGCAACAGCGCGCTAACCGATATTA 456

```

Db 421 GGTTTGGCAACGCGCTAACGATATTA 456

RESULT 6

AAc64625

ID AAC64625 standard; DNA; 456 BP.

AC AAC64625;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#4 DNA sequence SEQ ID NO:17.

XX Salmone11a: agfa: chromosomal gene replacement; fimbriin; epitope;

KW vaccine; immune response; immunogen; ds.

OS Salmone11a enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000MO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI: 2000-672631/65.

DR P-PSDB; AAB36349.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 136; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended

CC assembly system of strains of Salmone11a, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriin subunits, respectively; (2)

CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (3)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a Salmone11a, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (I) is

CC useful for the expression of recombinant Agfa protein which is useful for

CC eliciting an immune response in an animal. In a fimbriin presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbriin subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention

XX Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 79.6%; Score 363.2; DB 3; Length 456;

Best Local Similarity 87.3%; Pred.No. 7.3e-109;

Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGACAGCATTCGACGATCGTAGTTTCTGCGAGTCTTGACT 60

Db 1 ATGAACCTTTTAAAGTGGACAGCATTCGACGATCGTAGTTTCTGCGAGTCTTGACT 60

QY GCGCGTGTTCACATATGAGGCGGCGCGGTATATCAACGCGCGGCAATAGTCCGCG 120

Db 61 GCGCGTGTTCACATATGAGGCGGCGCGGTATATCAACGCGCGGCAATAGTCCGCG 120

QY 121 CCGGACTGACATGAGGATTTATGACGAGTTCGCTAACGCTGCGCTTCTGCA 180

Db 121 CCGGACTGATGATGAGCTGTATACCGGTGTGTATCCCATGAAATGCGACATGCA 180

QY 181 AGCGATGCGCGTAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240

Db 181 AGCGATGCGCGTAAATATGAAACGACATTCGACGAGCGGTATGATGATGATGATGAT 240

QY 241 GCGGCGGCGGTGCGGATATGATGATGATGATGATGATGATGATGATGATGATGAT 300

Db 241 GTAGGCGGAGGTGCGGATATGATGATGATGATGATGATGATGATGATGATGATGAT 300

QY 301 GCCACATGACACAGTGGAAAGCTTAAATCTCCATATTAATGTCGCGCAATACGCGGT 360

Db 301 GCCACATGACACAGTGGAAAGCTTAAATCTCCATATTAATGTCGCGCAATACGCGGT 360

QY 361 AATAACGCGCGCGTGTATATCAACGCGCATCTGATTCACGCTAATGTCGCTCAGTT 420

Db 361 AATAACGCGCGCGTGTATATCAACGCGCATCTGATTCACGCTAATGTCGCTCAGTT 420

QY 421 GGTTTGGCAACAGCGCGCTAACGATTTAA 456

Db 421 GGTTTGGCAACAGCGCGCTAACGATTTAA 456

RESULT 7

AAc64628

ID AAC64628 standard; DNA; 456 BP.

AC AAC64628;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#7 DNA sequence SEQ ID NO:23.

XX Salmone11a: agfa: chromosomal gene replacement; fimbriin; epitope;

KW vaccine; immune response; immunogen; ds.

OS Salmone11a enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000MO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI: 2000-672631/65.

DR P-PSDB; AAB36352.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
 Best Local Similarity 87.1%; Pred. No. 2.5e-108;

Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGAGCTCTGCT 60
 DB 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGAGCTCTGCT 60
 QY 61 GCGCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 DB 61 GCGCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 DB 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 QY 181 AGCGATGCCCGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 DB 181 AGCGATGCCCGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 240
 QY 241 GCAGGCGGAGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 241 GCAGGCGGAGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 301 GCCACCATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 DB 301 GCCACCATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 QY 361 AATTAACCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 DB 361 AATTAACCGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
 QY 421 GATTGGCAACAGCGCGCGGTAACTGATGATGATGATGATGATGATGATGATGAT 456
 DB 421 GATTGGCAACAGCGCGCGGTAACTGATGATGATGATGATGATGATGATGATGAT 456

RESULT 8
 ID AAC64622 standard; DNA; 456 BP.
 XX AAC64622;
 AC AAC64622;
 DT 26-FEB-2001 (first entry)
 XX

DE Agfa::PT3#1 DNA sequence SEQ ID NO:11.
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 XX WO200060102-A2.
 XX
 XX 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000MO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UUVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay WM;
 XX
 XX WPI: 2000-672631/65.
 XX
 XX P-PSDB: AAB36346.
 XX
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 XX Disclosure; Page 135; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
 Best Local Similarity 87.1%; Pred. No. 2.5e-108;
 Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGAGCTCTGCT 60
 DB 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGAGCTCTGCT 60
 QY 61 GCGCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 DB 61 GCGCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 DB 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180

DB 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCGCTTCTCTGCAA 180
QY 181 AGCGATGCGCGTAATATGATAGCTGTTACCCGTGTTTATCCCATGAATATGCAAT 240
DB 181 AGCGATGCGCGTAATATGATAGCTGTTACCCGTGTTTATCCCATGAATATGCAAT 240
QY 241 GCAGCGCCAGGTTGCGGATTAATAGTACTATTGTAAGTCACTCAAGATGTTTCAAGAAATAT 300
DB 241 GTAGGCGGAGGTTGCGGATTAATAGTACTATTGTAAGTCACTCAAGATGTTTCAAGAAATAT 300
QY 301 GCCACATCGACCACTGGAACGCTTAAACTCCGATATTACTGTCGCCAATACGCGCGT 360
DB 301 GCCACATCGACCACTGGAACGCTTAAACTCCGATATTACTGTCGCCAATACGCGCGT 360
QY 361 AATAACGCGCGGCTGTTAATCATACGCCATCTGATTCACAGCTTAATGTCGCTCAGGTT 420
DB 361 AATAACGCGCGGCTGTTAATCATACGCCATCTGATTCACAGCTTAATGTCGCTCAGGTT 420
QY 421 GGTTCGCAACAAACGCGACGCTTAACCATATTAA 456
DB 421 GCACATGCAACAAACGCGACGCTTAACCATATTAA 456
RESULT 9
ID AAC64623 standard; DNA; 456 BP.
AC AAC64623;
XX 26-FEB-2001 (first entry)
DT
XX
DE Agfa::PT#2 DNA sequence SEQ ID NO:13.
XX
KM Salmone11a; agfa, chromosomal gene replacement; fimbriin; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN MO20060102-A2.
XX
PD 12-OCT-2000.
XX
PP 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW,
XX WPI; 2000-672631/65.
DR P-PSDB; AAB36347.
XX
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139p; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF1/TA) nucleation depended
XX assembly system of strains of Salmone11a, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, Caga and Agfa-homologue fimbriin subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmone11a, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;
Query Match 79.3%; Score 361.6; DB 3; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.5e-108;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGCGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGCGAGTCTTGCT 60
QY 61 GCGCGCTTCGCAACATGGCGCGCGCGGCTTAATCATATAACGCGCGCAATAGTTCCGCGC 120
DB 61 GCGCGCTTCGCAACATGGCGCGCGCGGCTTAATCATATAACGCGCGCAATAGTTCCGCGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAAGCTGCGCTTCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAAGCTGCGCTTCTCTGCAA 180
QY 181 AGCGATGCGCGTAATATGATAGCTGTTACCCGTGTTTACCATGAATATGCAAT 240
DB 181 AGCGATGCGCGTAATATGATAGCTGTTACCCGTGTTTACCATGAATATGCAAT 240
QY 241 GCAGCGCCAGGTTGCGGATTAATAGTACTATTGTAAGTCACTCAAGATGTTTCAAGAAATAT 300
DB 241 GTAGGCGGAGGTTGCGGATTAATAGTACTATTGTAAGTCACTCAAGATGTTTCAAGAAATAT 300
QY 301 GCCACATCGACCACTGGAACGCTTAAACTCCGATATTACTGTCGCCAATACGCGCGT 360
DB 301 GCCACATCGACCACTGGAACGCTTAAACTCCGATATTACTGTCGCCAATATGATCAG 360
QY 361 AATAACGCGCGCTGTTAATCATACGCCATCTGATTCACAGCTTAATGTCGCTCAGGTT 420
DB 361 CTGCTTACCCGTGTTTACCATGAATATGCAATCGTAAAGCGTAATGTCGCTCAGGTT 420
QY 421 GGTTCGCAACAAACGCGACGCTTAACCATATTAA 456
DB 421 GGTTCGCAACAAACGCGACGCTTAACCATATTAA 456
RESULT 10
ID AAC64630 standard; DNA; 456 BP.
AC AAC64630;
XX 26-FEB-2001 (first entry)
DT
XX
DE Agfa::PT#9 DNA sequence SEQ ID NO:27.
XX
KM Salmone11a; agfa, chromosomal gene replacement; fimbriin; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN MO20060102-A2.
XX

PD 12-OCT-2000.
 XX 05-APR-2000; 2000OWO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay MW;
 XX WPI; 2000-672631/65.
 DR P-PSDB; AAB36354.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;
 Query Match 78.9%; Score 360; DB 3; Length 456;
 Best Local Similarity 86.8%; Pred. No. 8.3e-108;
 Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 1 ATGAACTTTTAAAGTGGAGCATTCGACGATTCGAGTTCGTGAGTGGCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTCGACGATTCGAGTTCGTGAGTGGCTGGCT 60
 QY 61 GGGCGCTTCCACATGAGGGGGGGGGGGGATATATACGCGCGCAATAGTTCGGCG 120
 DB 61 GGGCGCTTCCACATGAGGGGGGGGGGGGATATATACGCGCGCAATAGTTCGGCG 120
 QY 121 CCGAGCTCAACGTTGAGCATTTACGATGAGTTCGCTGACGCTGCTGCTGCA 180
 DB 121 CCGAGCTCAACGTTGAGCATTTACGATGAGTTCGCTGACGCTGCTGCTGCA 180
 QY 181 AGCGATCCCGGTAAATATGATCAGCTGCTTACCCGATGATGCAATGAGCAGAT 240
 DB 181 AGCGATCCCGGTAAATATGATCAGCTGCTTACCCGATGATGCAATGAGCAGAT 240
 QY 241 GCAGGCGAGGTGGGATATAGTACATGAGTCAAGTCAAGTGTTCGAAATAT 300
 DB 241 TATGATCAGCTGATCCGCTGTTGTTACCAAGAAATGCAATGCAATTCGAAATAT 300
 QY 301 GCCACATGACGACGATGGAACGCTAAACCTCGATATTAATCTGTGGCGCAATAGCGCGGT 360

DB 301 GCCACATGACGACGATGGAACGCTAAACCTCGATATTAATCTGTGGCGCAATAGCGCGGT 360
 QY 361 AATAACGCGCGGTGTTAAATCAGACCGCATTCGATTCAGCGTAAATGGGCGTCAAGTT 420
 DB 361 AATAACGCGCGGTGTTAAATCAGACCGCATTCGATTCAGCGTAAATGGGCGTCAAGTT 420
 QY 421 GGTTCGCAACAGCGACGCGTAAACGATTTAA 456
 DB 421 GGTTCGCAACAGCGACGCGTAAACGATTTAA 456
 RESULT 11
 AAC64624
 ID AAC64624 standard; DNA; 456 BP.
 XX
 AC AAC64624;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#3 DNA sequence SEQ ID NO:15.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN WC200060102-A2.
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000OWO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay MW;
 XX WPI; 2000-672631/65.
 DR P-PSDB; AAB36348.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SO Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 78.9%; Score 360; DB 3; Length 456;

Best Local Similarity 86.8%; Pred. No. 8.3e-108;

Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
OY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTGGCT 60
OY 61 GGGCGTGTTCACATGGGGCGGCGGTATCTAATACGGCGGCAATAGTTCCGGC 120
DB 61 GGGCGTGTTCACATGGGGCGGCGGTATCTAATACGGCGGCAATAGTTCCGGC 120
OY 121 CCGGACTCAAGCTTACGATTTATCAGTACGGTTCCGCTAACGGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTACGATTTATCAGTACGGTTCCGCTAACGGCTGCTCTGCA 180
OY 181 AGCGATGCCGTAAATATGATCAGTGGTTACCGGTGTTTACCCATGAATGGCAG 240
DB 181 AGCGATGCCGTAAATATGATCAGTGGTTACCGGTGTTTACCCATGAATGGCAG 240
OY 241 GCAGGCGGAGTGGCGGATTAATAGTACTATGACATCTCAGATGTTTCAAGAAATAT 300
DB 241 GCAGGCGGAGTGGCGGATTAATAGTACTATGACATCTCAGATGTTTCAAGAAATAT 300
OY 301 GCCACATCGACGATGGAACGCTTAAACTCCGATATTAATGTCGGGCAATACGGCG 360
DB 301 GCCACATCGACGATGGAACGCTTAAACTCCGATATTAATGTCGGGCAATACGGCG 360
OY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGG 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGG 420
OY 421 GGTGTTGGCAACACGCGGCTTAAACGATTTAA 456
DB 421 GGTGTTGGCAACACGCGGCTTAAACGATTTAA 456
```

RESULT 12

AAC64627
ID AAC64627 standard; DNA; 456 BP.

XX AAC64627;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#6 DNA sequence SEQ ID NO:21.

XX Salmonella: agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX MO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.

DR P-PSDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP17/TA) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SO Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 78.6%; Score 358.4; DB 3; Length 456;

Best Local Similarity 86.6%; Pred. No. 2.8e-107;

Matches 395; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```
OY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTGGCT 60
OY 61 GGGCGTGTTCACATGGGGCGGCGGTATCTAATACGGCGGCAATAGTTCCGGC 120
DB 61 GGGCGTGTTCACATGGGGCGGCGGTATCTAATACGGCGGCAATAGTTCCGGC 120
OY 121 CCGGACTCAAGCTTACGATTTATCAGTACGGTTCCGCTAACGGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTACGATTTATCAGTACGGTTCCGCTAACGGCTGCTCTGCA 180
OY 181 AGCGATGCCGTAAATATGATCAGTGGTTACCGGTGTTTACCCATGAATGGCAG 240
DB 181 AGCGATGCCGTAAATATGATCAGTGGTTACCGGTGTTTACCCATGAATGGCAG 240
OY 241 GCAGGCGGAGTGGCGGATTAATAGTACTATGACATCTCAGATGTTTCAAGAAATAT 300
DB 241 GCAGGCGGAGTGGCGGATTAATAGTACTATGACATCTCAGATGTTTCAAGAAATAT 300
OY 301 GCCACATCGACGATGGAACGCTTAAACTCCGATATTAATGTCGGGCAATACGGCG 360
DB 301 GCCACATCGACGATGGAACGCTTAAACTCCGATATTAATGTCGGGCAATACGGCG 360
OY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGG 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGG 420
OY 421 GGTGTTGGCAACACGCGGCTTAAACGATTTAA 456
DB 421 GGTGTTGGCAACACGCGGCTTAAACGATTTAA 456
```


PA (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
PI Kay WW, Collinson SK, Clouthier SC, Doran JL,
XX WPI; 1994-358275/44.
DR P-PSDB; AAR62761.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compans. contg. fimbrial type proteins.
XX
PS Disclosure; Fig 7A; 95pp; English.
XX
CC The DNA encodes the Salmonella enteritidis27655-3b TnpH α mutant strain
CC agfa gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to Salmonella in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 63.3%; Score 288.6; DB 2; Length 361;
Best Local Similarity 91.3%; Pred. No. 2.3e-84;
Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 64 GTCGTTCCACATGGGGCGGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCCCG 123
DB 1 GTCGTACACAGATGGGCGGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCCCG 60

QY 124 GACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAAGCGGCGCTTCTGTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAAGCGGCGCTTCTGTGCAAGC 120

QY 184 GATGCCCGTAAATATGATCAGCTGTGTTACCCGTGTTTACCAATGACCAATGCA 243
DB 121 GATGCCCGTAAATATGATCAGCTGTGTTACCCGTGTTTACCAATGACCAATGCA 180

QY 244 GGCAGGGGTGGGATTAATGATTAATTAAGTGAAGTGAATGTTTCAAAATTAATGCC 303
DB 181 GGCAGGGGTGGGATTAATGATTAATTAAGTGAAGTGAATGTTTCAAAATTAATGCC 240

QY 304 ACCATGACGAGTGAACGCTAAACCTCCGATTAATGTCGGCCAAATACGGCGGTAT 363
DB 241 ACCATGACGAGTGAACGCTAAACCTCCGATTAATGTCGGCCAAATACGGCGGTAT 300

QY 364 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 335

RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
AC AAT74141;
XX
DT 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
DE Salmonella enteritidis 27655-3b TnpH α mutant agfa gene fragment.
XX
KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; db.
XX
XX Salmonella enteritidis.
OS
FH Key location/Qualifiers
FT CDS 1..360
FT /tag= a
FT /label= agfa_gene_fragment
FT primer_bind 16..60
FT /tag= b
FT /label= Primer_TAF5

FT primer_bind 52..69
FT /tag= c
FT /label= Primer_TAF3
FT complement(103..128)
FT primer_bind
FT /tag= d
FT /label= Primer_TAF6
FT complement(294..312)
FT /tag= e
FT /label= Primer_TAF4
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX
XX WPI; 1997-309886/28.
XX
XX P-PSDB; AAW23569.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX
XX Claim 1; Col 107-110; 85pp; English.
XX
XX The present sequence represents an isolated agfa gene fragment derived
XX from Salmonella enteritidis 27655-3b TnpH α mutant strain. The nucleic
XX acid can be used to provide diagnostic assays for Salmonella and/or
XX enteropathogenic bacteria of the family Enterobacteria. It can also be
XX used to provide proteins and antibodies which can be used for assays. The
XX nucleic acid sequence can be used to provide probes or primers which can
XX specifically hybridise to nucleic acid molecules from greater than 99% of
XX Salmonella strains that are pathogenic to warm-blooded animals relative
XX to nucleic acid molecules from virtually all other microbial organisms.
XX (Updated on 25-MAR-2003 to correct PF field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 63.3%; Score 288.6; DB 2; Length 361;
Best Local Similarity 91.3%; Pred. No. 2.3e-84;
Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 64 GTCGTTCCACATGGGGCGGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCCCG 123
DB 1 GTCGTACACAGATGGGCGGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCCCG 60

QY 124 GACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAAGCGGCGCTTCTGTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATCAGTACGCTCCGCTAAGCGGCGCTTCTGTGCAAGC 120

QY 184 GATGCCCGTAAATATGATCAGCTGTGTTACCCGTGTTTACCAATGACCAATGCA 243
DB 121 GATGCCCGTAAATATGATCAGCTGTGTTACCCGTGTTTACCAATGACCAATGCA 180

QY 244 GGCAGGGGTGGGATTAATGATTAATTAAGTGAAGTGAATGTTTCAAAATTAATGCC 303
DB 181 GGCAGGGGTGGGATTAATGATTAATTAAGTGAAGTGAATGTTTCAAAATTAATGCC 240

QY 304 ACCATGACGAGTGAACGCTAAACCTCCGATTAATGTCGGCCAAATACGGCGGTAT 363
DB 241 ACCATGACGAGTGAACGCTAAACCTCCGATTAATGTCGGCCAAATACGGCGGTAT 300

QY 364 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 335

Search completed: March 15, 2004, 17:52:03

Thu Mar 18 12:28:02 2004

Job time : 246.684 secs

us-09-543-407-19.rng

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OW nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds

(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456
Sequence: 1 atgaacttcaaaagtcgc.....ccacgctcaaccagratraa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

1: Issued Patents NA: *
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/6C.COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412.8	90.5	456	1	US-08-233-788A-58
2	288.6	63.3	361	1	US-08-233-788A-56
3	34	7.5	1344	3	US-09-120-927-1
4	34	7.5	1344	4	US-09-431-614-5
5	32.8	7.2	1218	4	US-09-489-039A-403
6	32.2	7.1	399	4	US-09-621-976-8976
7	32	7.0	1152	4	US-09-489-039A-2755
8	31.4	6.9	534	4	US-09-489-039A-2195
9	31.4	6.9	1818	4	US-09-489-039A-2162
10	31.2	6.8	549	4	US-09-252-991A-14225
11	31.2	6.8	696	4	US-09-252-991A-14137
12	31	6.8	2052	1	US-08-443-104-5
13	31	6.8	2052	1	US-08-238-130-6
14	31	6.8	2052	1	US-08-442-859-5
15	31	6.8	2052	2	US-08-398-489-5
16	31	6.8	2052	2	US-08-894-772-1
17	31	6.8	2052	2	US-09-207-844-1
18	31	6.8	2052	4	US-09-252-509-1
19	31	6.8	2052	5	PCT-US95-05534-5
20	30.8	6.8	1662	1	US-08-565-386-2
21	30.8	6.8	2547	3	US-08-508-761B-1
22	30.8	6.8	4529	1	US-08-565-386-1
23	30.8	6.8	1664976	4	US-08-916-421B-1
24	30.6	6.7	4403765	3	US-09-103-840A-2
25	30.6	6.7	4411529	3	US-09-103-840A-1
26	30	6.6	2436	4	US-09-540-236-492
27	30	6.6	3895	4	US-08-961-527-201

C	28	30	6.6	65792	4	US-09-596-002-31	Sequence 31, Appl
C	29	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
C	30	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
C	31	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
C	32	29.8	6.5	51259	3	US-08-781-891A-209	Sequence 209, App
C	33	29.8	6.5	51259	4	US-09-618-166-209	Sequence 209, App
C	34	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
C	35	29.4	6.4	912	4	US-09-543-681A-2040	Sequence 2040, Ap
C	36	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
C	37	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
C	38	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
C	39	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appl
C	40	29.2	6.4	708	4	US-09-543-681A-2551	Sequence 2551, Ap
C	41	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C	42	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C	43	29	6.4	939	4	US-09-976-594-108	Sequence 108, App
C	44	29	6.4	1097	4	US-09-988-784-7	Sequence 7, Appl
C	45	29	6.4	98844	4	US-09-791-211-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Dorian, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 90.5%; Score 412.8; DB 1; Length 456;
Best Local Similarity 94.1%; Pred. No. 1.7e-137;
Matches 429; Conservative 0; Mismatches 27; Indels 0;

QY 1 ATGAACCTTTTAAAGTGAGCAATTCCAGCAATCGATTCTGCGAGTCTGCTGCT 60
 Db 1 ATGAACCTTTTAAAGTGAGCAATTCCAGCAATCGATTCTGCGAGTCTGCTGCT 60
 QY 61 GCGCGTCTTCCACATAGGCGGCGGCGGCTATCATACCGCGCGCAATAGTTCGGC 120
 Db 61 GCGCGTCTTCCACATAGGCGGCGGCGGCTATCATACCGCGCGCAATAGTTCGGC 120
 QY 121 CCGGACTCAACGTTGAGCAATTTATCAATACGTTCCGCTAAGCGTCCGCTCTGCA 180
 Db 121 CCGGACTCAACGTTGAGCAATTTATCAATACGTTCCGCTAAGCGTCCGCTCTGCA 180
 QY 181 AGCGATCCCGTAAATATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACAT 240
 Db 181 AGCGATCCCGTAAATATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACAT 240
 QY 241 GCGGCGCGGCTCGGATTAATAGTACTATTAATGATGATGATGATGATGATGAT 300
 Db 241 GCGGCGCGGCTCGGATTAATAGTACTATTAATGATGATGATGATGATGATGAT 300
 QY 301 GCCACCATCGACAGGAGAAAGCTTAAATCTCGATTTACTGTCGCGCAATAGCGCGT 360
 Db 301 GCCACCATCGACAGGAGAAAGCTTAAATCTCGATTTACTGTCGCGCAATAGCGCGT 360
 QY 361 AATAAGCGCGCTGCTGTTATATCAGACCGCATCTGATTCGAGCGTATGCTGATG 420
 Db 361 AATAAGCGCGCTGCTGTTATATCAGACCGCATCTGATTCGAGCGTATGCTGATG 420
 QY 421 GGTTTTGGCAACAGCGCAAGCGCTTACCGATTTAA 456
 Db 421 GGTTTTGGCAACAGCGCAAGCGCTTACCGATTTAA 456

RESULT 2

US-08-233-788A-56
 ; Sequence 56, Application US/08233788A
 ; Patent No. 3635617
 ; GENERAL INFORMATION:

APPLICANT: Doran, James L.
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, Karen S.
 APPLICANT: Clouthier, Sharon C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 NUMBER OF SEQUENCES: 61
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043, 403C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 682-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEDANBERRY
 INFORMATION FOR SEQ ID NO: 56:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 base pairs

TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..357
 US-08-233-788A-56

Query Match

Best Local Similarity 63.3%; Score 288.6; DB 1; Length 361;
 Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 64 GTCTTCCACATAGGCGGCGGCGGCTATCATACGCGCGGCAATAGTTCGGCCCG 123
 Db 1 GTCTTCCACATAGGCGGCGGCGGCTATCATACGCGCGGCAATAGTTCGGCCCG 123
 QY 124 GACTCAACGTTGACATTTATCAGTACGCTTCCGCTAAGCGTCCGCTCTGCAAGC 183
 Db 61 GACTCAACGTTGACATTTATCAGTACGCTTCCGCTAAGCGTCCGCTCTGCAAGC 120
 QY 184 GATGCCCGTAATATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACATGCA 243
 Db 121 GATGCCCGTAATATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACATGCA 180
 QY 244 GCGGCGGCTCGGATTAATAGTACTATTAATGATGATGATGATGATGATGATG 303
 Db 181 GCGGCGGCTCGGATTAATAGTACTATTAATGATGATGATGATGATGATGATG 240
 QY 304 ACCATCGACAGTGAAGCGTAAATCTCGATTTACTGTCGCGCAATAGCGCGTAA 363
 Db 241 ACCATCGACAGTGAAGCGTAAATCTCGATTTACTGTCGCGCAATAGCGCGTAA 300
 QY 364 AACGCGCGCTGCTGTTATATCAGACCGCATCTGATTC 398
 Db 301 AACGCGCGCTGCTGTTATATCAGACCGCATCTGATTC 335

RESULT 3

US-09-120-927-1/c
 ; Sequence 1, Application US/09120927
 ; Patent No. 6262018
 ; GENERAL INFORMATION:

APPLICANT: Kim, Jihyun Francis
 APPLICANT: Beer, Steven V.
 TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
 STREET: P.O. Box 1051, Clinton Square
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/120,927
 FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/055,108
 FILING DATE: 06-AUG-1977
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldmar, Michael L.
 REGISTRATION NUMBER: 30,727
 REFERENCE/DOCKET NUMBER: 19603/1581
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1344 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-120-927-1

Query Match
Best Local Similarity 54.9%; Score 34; DB 3; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGGCGTGTTCACAAATGGGCGCGCGGTAATCATACGCGCGGCAT 111
DB 209 GCTCGGTTGCGCGCATTCGATGATGCGCGATACAGACGTTTACGCTCCGCAT 150
QY 112 AGTTCGCGCGCGCATTCAGATTATTCAGTTCGCTTACGCTGCGCTT 171
DB 149 AATTGACCATTTGCTCATGTTGCGCATGATGCGTGTGCCCCACAGCAATTT 90
QY 172 GC 173
DB 89 GC 88

RESULT 4
US-09-431-614-5/c
Sequence 5, Application US/09431614
Patent No. 6624139
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
FILE REFERENCE: 21829/41 (EBC-003)
CURRENT APPLICATION NUMBER: US/09/431,614
EARLIER FILING DATE: 1999-11-02
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-431-614-5

Query Match
Best Local Similarity 7.5%; Score 34; DB 4; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGGCGTGTTCACAAATGGGCGCGCGGTAATCATACGCGCGGCAT 111
DB 209 GCTCGGTTGCGCGCATTCGATGATGCGCGATACAGACGTTTACGCTCCGCAT 150
QY 112 AGTTCGCGCGCGCATTCAGATTATTCAGTTCGCTTACGCTGCGCTT 171
DB 149 AATTGACCATTTGCTCATGTTGCGCATGATGCGTGTGCCCCACAGCAATTT 90
QY 172 GC 173
DB 89 GC 88

RESULT 5
US-09-489-039A-403
Sequence 403, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709,2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 403
LENGTH: 1218
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-403

Query Match
Best Local Similarity 58.0%; Score 32.8; DB 4; Length 1218;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 174 TCTGCAAGCGATCCCGTAATATGATCAGCTGTTACCGTGTGTACCAATGAAT 233
DB 614 TCGTCAAGCGCTGATCTCTCATATGTTACAGACCTTATGCTGTGCGTAATACC 673
QY 234 GGCACATGACGCGCGGCGGATGATGATGATGATGATGATGATGATGATGAT 273
DB 674 CGACG 713

RESULT 6
US-09-621-976-8976/c
Sequence 8976, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jodert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent .pm
SEQ ID NO 8976
LENGTH: 399
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match
Best Local Similarity 7.1%; Score 32.2; DB 4; Length 399;
Matches 34; Conservative 151; Mismatches 154; Indels 0; Gaps 0;

QY 88 GGTATCATTAACGCGCGCGCGCAATAGTTCGCGCGCGCATCACTGATGACATTTATCAG 147
DB 340 GMYTMYMTSRRGSGYRKYTSAMMGRAMKMKKTKMYMKGGKGGSTYMMRSRG 281
QY 148 TAGCGTCCGTAACGCGCGCGCTGCTGCAAGGATGCGCGGTAATATGATCAGCTG 207
DB 280 STGRWSTRRAMWRSKSGSYRMYRSMYSAMWRRKMTCTWGRSGSRS 221
QY 208 GTTACCGCTGTTGTTACCATGAATGACACATGACGCGCGGCGGATGATGATGAT 267
DB 220 TGYAAMWYKSKWCKSRKMYKRRKRCSTSTCTGSGSYKCKAYTTKKRRWT 161
QY 268 ATTGAAGTACTGAGAAATGTTTCAAGAAATATCCACATGACGACGAGTGAACGCTAAA 327
DB 160 RTYVY 101
QY 328 AACTCCGATATTAATGTCGCGCATATGCGGCGGTAATACCGCGCGGCTGTTAATCAAGCC 387
DB 100 YWYVY 41
QY 388 GCATCTGATTCACAGCGTATGTCGTCAGGCTGCTGCTGCTGCTGCTGCTGCTGCT 426
DB 40 WSKWTYVY 2

Query Match	7.0%;	Score 32;	DB 4;	Length 1152;
Best Local Similarity	56.7%;	Pred. No. 0.43;		
Matches 59;	Conservative	0;	Mismatches	45.
				Total

Oy	155	CGCGTAAAGCGTGCCTTGTCTGTGCAAGACGATGCCCGTAATATGATGACAGTGGTAAACC	214
Db	656	CCACGGCGCTGGGCTGGCTGTCTGTCTCAAGCGGCTGCTGATAGTACACAGTGGCGATG	715
Oy	215	GTGTTTATCCCATTAATGCGACATGCAAGCCAGAGTGGCGAT	258
Db	716	TGCTGATTCCCATTAATGACCGCGTGGCGACAGGCGCGGAT	759

```

RESULT 8
US-09-489-039A-2195/C
; Sequence 2195 Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709,2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2195
; LENGTH: 534
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2195

```

Query Match	6.9%	Score 31.4	DB 4	Length 534
Best Local Similarity	52.7%	Pred. No. 0	44	
Matches	68	Conservative	0	Mismatches 61
				Indels 0
				Gaps 0
QY	74	AATGGGCGGCGCGGTAATCATTAACGGCGGCGCAATAGTCTCCGCGCGGACTCAAGCT	133	
Db	501	AAAGGCGCGAGCGGCGGTACCGGTGCTGACCGCTGAGGAGCGAATCAAGGCG	442	
QY	134	TGAGCATTTATCAGTACGCTTCCGCTAACGCTGCGCTTCTCTGGCAAGCGATGCCGGA	193	
Db	441	GCGCTTTTAGCGCGCTGCGTTACGGCGCTGTGACGCGCTTTCTGGAAACGATCCCGGTT	382	
QY	194	AATATATC 202		
Db	381	TATATCTC 373		

RESULT 9

```

US-09-489-039A-2162
/ Sequence 2162: Application US/09489039A
/ Patent No. 6610836
/ GENERAL INFORMATION:
/ APPLICANT: Gary Breton et. al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/ FILE REFERENCE: 2709.2004001
/ CURRENT APPLICATION NUMBER: US/09/489,039A
/ PRIORITY FILING DATE: 2000-01-27
/ PRIOR APPLICATION NUMBER: US 60/117,747
/ PRIORITY FILING DATE: 1999-01-29
/ NUMBER OF SEQ ID NOS: 14342
/ SEQ ID NO 2162
/ LENGTH: 1818
/ TYPE: DNA
/ ORGANISM: Klebsiella pneumoniae
/ FEATURE:
/ NAME/KEY: unsure
/ LOCATION: (702)
/ OTHER INFORMATION: Identity of nucleotide at the above locations are unknown.
US-09-489-039A-2162

```

Query Match	6.9%;	Score 31.4;	DB 4;	Length 1818,
Best Local Similarity	52.7%;	Pred. No. 0.92;		
Matches 68;	Conservative	0;	Mismatches	21

	Sequences	61	Indels	0	Gaps	0
QY	74	AATGGGGCGGGCGGTTATCATTAACGGCGCGGCATATAGTTCCGCGCCGGACTAAAGT				
Db	1363	AAAGAGCGCGGCACCGACCGCGTGGCGGTGCTGTGCACCGCTGGAGCACTAGGGC				
QY	134	TGAGCATTTATCAGTACGTTCCGCTAACGCTGCGTCTCTGTGCAAGGCAATGCCGTA				1422
Db	1423	GGCTTCTTTACCGCGCTGAGGCGGCTGTGACGCTGTCTGGAACCAACGATCCCGTT				1493
QY	194	AATATGATGC	202			
Db	1483	TATATCTTC	1491			

```

RESULT 10
US-09-252-991A-14225/c
; Sequence 14225: Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14225
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14225

```

Query Match	6.8%;	Score 31.2;	DB 4;	Length 545;
Best Local Similarity	53.2%;	Pred. No. 0.53;		
Matches	66;	Conservative	0;	Mismatches 58; Indels 0; Gaps 0
Oy	131	CGTTGACATTATCAGTACGGTCCGCTAACGCTGGCGCTTCTCTGCACAAAGCATGCC		190
Db	282	CGTTGACCGGTGAGGCTCGGGGTGAGGCCGGTGTGCATCGATCTGGATATGCGCGCTG		223
Oy	191	GTAAATATATAGTGGTTTACCCGTGTTTAAACCATGAAATGAGCAGATGACGAGCGAGG		250
Db	222	GTGCCGAGGGTCTGGCGGTGCACCATCTTGTAAGTCTCGAATAGGTGTGCCGAATCTCAGG		163

QY 251 GTGC 254
Db 162 GTTC 159

RESULT 11

US-09-252-991A-14137/C
Sequence 14137, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rudenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14137
LENGTH: 696
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14137

Query Match 6.8%; Score 31.2; DB 4; Length 696;
Best Local Similarity 53.2%; Pred. No. 0.61;
Matches 66; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 131 CGTTGACGATTATTCAGTACGCTTCCCTAACGCTGCGCTTGTCTGCAAGCGATGCC 190
Db 377 CGTTGACGCTTACGCTGCGCTTCCCTAACGCTGCGCTTGTCTGCAAGCGATGCC 318
QY 191 GTAATATGATCAGCTGCTTACCGCTGTTGTTACCCATGAATGGCAGATGCGCCAGG 250
Db 317 GTGCGGACGCTGCTGCGCTTACCGCTGTTGTTACCCATGAATGGCAGATGCGCCAGG 258
QY 251 GTGC 254
Db 257 GTTC 254

RESULT 12

US-08-443-104-5/C
Sequence 5, Application US/08443104
Patent No. 5691162
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fugleang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5691162 No. 5691162disk of No. 5691162th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,104
FILING DATE: 17-MAY-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-9655
TELEFAX: 212-867-0123
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-443-104-5

Query Match 6.8%; Score 31; DB 1; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 215 GTGTTTACCCAGTAATGCGACATGCGCCAGGGTGGCGATTAATGACTATTGAAAC 274
Db 648 GTGAGCTTACGCTGCGGCTTGTGACGTAACCTTGTGACGTAAGAGGTGATGTTTC 589
QY 275 TGACTGAAATGTTTGAAGAAATATGACCATGACAGTGAAGCGTTAAACTCCG 334
Db 588 TTGCTCTTCTTACGCTTGTGACCTTGTGACCTTGAACCGGAGAGAAACATCAAG 529
QY 335 ATATTACTGTGCGCAATACGCGGTATTAACGCGGCTGTTTATCAACGCGATCTG 394
Db 528 GTGCTCTTGAAGAGCGTCAACGCGGCTTCTCTGTCAGCTGTAAAGAGACCGGAATC 469
QY 395 ATTCCAGCTTAATGTCGCTGCTGTTTGGCAACACGCAACGCTTAACAGTAT 453
Db 468 TTGCTCTTGAAGAGCTGTTTCTGTAAGAGACCTGCGCTGACGCGCAATCTGAT 410

RESULT 13

US-08-238-130-6/C
Sequence 6, Application US/08238130
Patent No. 5702934
GENERAL INFORMATION:
APPLICANT: Hastrup, Sven
APPLICANT: Branner, Sven
APPLICANT: Jorgensen, Birthe R.
APPLICANT: Christensen, Tove
APPLICANT: Jorgensen, Birgitte B.
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fugleang, Claus
TITLE OF INVENTION: PROCESSES FOR PRODUCING AN ENZYME
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5702934 No. 5702934disk of No. 5702934th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,130
FILING DATE: 04-MAY-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: DK 522/93
FILING DATE: 05-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 3965, 200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-9655
INFORMATION FOR SEQ ID NO: 6
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..363, 416..802, 856..1821, 1870..2052)
US-08-442-859-6

Query Match
Best Local Similarity 6.8%; Score 31; DB 1; Length 2052;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 215 GTGTTGTTACCCATGAAATGACATGACGAGCGGATGATATGACTATTGAAAC 274
DB 648 GTGAGCTTACCTCGGGCTTGCTGACGTAACCTTGTGTAACAAGGTAGAGGTTTC 589
QY 275 TGACTGAAATGTTTCAGAAATATGACCATGACCATGGAACGCTTAAACCTCCG 334
DB 588 TTGCTCTTCTTTCAGCCCTTGCGCTGTGACGCTCAAGGGAGAAAGAACTCAACG 529
QY 335 ATATTACTGTGCGCCATATGCGGGTAATACGCGCGCTGTTATATGACCGCATCTG 394
DB 528 GTGTCCTTAGAGCGTCGACGGGGCTTCTGTCACGCTGTGTAAGAGACCGGAATC 469
QY 395 ATTCCAGGCTAATGTGCGTCAAGTTGTTTGGCAACACGCGCTAACCAAGTAT 453
DB 468 TTGCCCTCTAGAGAGCTGTTTCCGTAGAGAAAGACCTCGCCCTCAGCGCAATCTGAT 410

RESULT 14
US-08-442-859-5/C
Sequence 5, Application US/08442859
Patent No. 5807729
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fugleang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5807729 No. 5807729disk of No. 5807729th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,859

FILING DATE: 17-MAY-1995
CLASSIFICATION: 252
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180, 010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-867-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-442-859-5

Query Match
Best Local Similarity 6.8%; Score 31; DB 1; Length 2052;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 215 GTGTTGTTACCCATGAAATGACATGACGAGCGGATGATATGACTATTGAAAC 274
DB 648 GTGAGCTTACCTCGGGCTTGCTGACGTAACCTTGTGTAACAAGGTAGAGGTTTC 589
QY 275 TGACTGAAATGTTTCAGAAATATGACCATGACCATGGAACGCTTAAACCTCCG 334
DB 588 TTGCTCTTCTTTCAGCCCTTGCGCTGTGACGCTCAAGGGAGAAAGAACTCAACG 529
QY 335 ATATTACTGTGCGCCATATGCGGGTAATACGCGCGCTGTTATATGACCGCATCTG 394
DB 528 GTGTCCTTAGAGCGTCGACGGGGCTTCTGTCACGCTGTGTAAGAGACCGGAATC 469
QY 395 ATTCCAGGCTAATGTGCGTCAAGTTGTTTGGCAACACGCGCTAACCAAGTAT 453
DB 468 TTGCCCTCTAGAGAGCTGTTTCCGTAGAGAAAGACCTCGCCCTCAGCGCAATCTGAT 410

RESULT 15
US-08-398-489-5/C
Sequence 5, Application US/08398489
Patent No. 5843753
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fugleang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5843753 No. 5843753disk of No. 5843753th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 03-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.

REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-398-489-5

Query Match 6.8%; Score 31; DB 2; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4; Mismatches 130; Indels 0; Gaps 0;
Matches 109; Conservative 0;

Oy	215	GTGTTGTTACCATGAATGCGACATGCGAGGCGGATATAGTACTATTGAAC	274
Db	648	GTGAGCTTAGCCTCGGGCTTGCTGACGCTTGTGTAACGTAAGTGTAGTGTTC	589
Oy	275	TGACTCAGAAATGTTTCAGAAATAATGCCACCATGACCAATGAACGCTAAAACTCCG	334
Db	588	TTGCTCTTCTCTCAGCCTTGCGCTTGTCAAGCTCAACGGGAGAGAAAGAACATCAACG	529
Oy	335	ATATTACTGTGGCCCAATACGGCGGTAAATACGCCCGCTGTTATCAGACCGCATCTG	394
Db	528	GTGTCCTTGAGAGCGTCGACGGGGCTTCTGTCACGCTTGTTAAGAGAACCGGGAATC	469
Oy	395	ATTCCAGCGTATGCTGCGTCAGTTGGTTTGGCAACGCCACGGCTAACCGATAT	453
Db	468	TTGCCCTGTAGAAAGCTGTTCCGTAGAGAAAGACCTGCCGCTCAGCGCAATCTGTAT	410

Search completed: March 16, 2004, 04:37:54
Job time: 58.6647 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query	Length	DB	ID	Description
c 1	34.6	7.6	1662	14	US-10-156-761-1483	Sequence 3483, App
c 2	34.6	7.6	9025608	14	US-10-156-761-1	Sequence 1, Appl
c 3	34	7.5	1344	9	US-09-833-681-6	Sequence 6, Appl
c 4	34	7.5	1344	9	US-09-880-371-6	Sequence 6, Appl
c 5	34	7.5	1344	9	US-09-879-248-5	Sequence 5, Appl
c 6	34	7.5	1344	14	US-10-010-390-6	Sequence 6, Appl
c 7	34	7.5	1344	15	US-10-441-726-5	Sequence 5, Appl
c 8	33.4	7.3	655	14	US-10-184-634-118	Sequence 418, Appl
c 9	33.4	7.3	655	14	US-10-184-634-118	Sequence 418, App
c 10	33.2	7.3	1848	14	US-10-156-761-7222	Sequence 7222, App
c 11	33.2	7.3	9025608	14	US-10-156-761-1	Sequence 1, Appl
c 12	32.8	7.2	24081	14	US-10-132-134-13	Sequence 13, Appl
c 13	32.8	7.2	52101	14	US-10-132-134-13	Sequence 1, Appl
c 14	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, App
c 15	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, App

C	23	31.2	6.8	6.8	2266	15	US-10-156-761-687	Sequence 687, App
C	22	31.4	6.9	7374	14	US-10-505-428-1006	Sequence 10067, A	
C	21	31.4	6.9	7374	10	US-09-764-891-10207	Sequence 10207, A	
C	20	31.4	6.9	460	15	US-10-260-238-787	Sequence 787, App	
C	19	31.6	7.0	6003	13	US-10-010-901-9	Sequence 9, App1	
C	18	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A	
C	17	31.8	7.0	930	12	US-10-369-493-24480	Sequence 24480, A	
C	16	31.8	7.0	930	15	US-10-382-122A-20335	Sequence 20335, A	
C	15	31.8	7.0	930	12	US-10-382-122A-20430	Sequence 20430, A	
C	14	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A	
C	13	31.6	7.0	978	15	US-10-369-493-24265	Sequence 24265, A	
C	12	31.4	6.9	460	15	US-10-260-238-787	Sequence 787, App	
C	11	31.4	6.9	7374	10	US-09-764-891-10207	Sequence 10207, A	
C	10	31.4	6.9	7374	14	US-10-505-428-1006	Sequence 10067, A	
C	9	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App	
C	8	31.2	6.8	2266	15	US-10-108-260A-375	Sequence 375, App	
C	7	31.1	6.8	1049	14	US-10-123-155-358	Sequence 358, App	
C	6	31.1	6.8	1049	14	US-10-146-731-358	Sequence 358, App	
C	5	31.1	6.8	1049	14	US-10-140-472-358	Sequence 358, App	
C	4	31.1	6.8	1049	14	US-10-141-761-358	Sequence 358, App	
C	3	31.1	6.8	1049	14	US-10-142-883-358	Sequence 358, App	
C	2	31.1	6.8	1049	14	US-10-158-790-358	Sequence 358, App	
C	1	31.1	6.8	1049	15	US-10-137-871-358	Sequence 358, App	
C	0	31.1	6.8	1049	15	US-10-140-923-358	Sequence 358, App	
C	33	31.1	6.8	1049	15	US-10-141-756-358	Sequence 358, App	
C	32	31.1	6.8	1049	15	US-10-140-923-358	Sequence 358, App	
C	31	31.1	6.8	1049	15	US-10-140-805-358	Sequence 358, App	
C	30	31.1	6.8	1049	15	US-10-140-805-358	Sequence 358, App	
C	29	31.1	6.8	1049	15	US-10-141-756-358	Sequence 358, App	
C	28	31.1	6.8	1049	14	US-10-141-761-358	Sequence 358, App	
C	27	31.1	6.8	1049	14	US-10-140-472-358	Sequence 358, App	
C	26	31.1	6.8	1049	14	US-10-146-731-358	Sequence 358, App	
C	25	31.1	6.8	1049	14	US-10-123-155-358	Sequence 358, App	
C	24	31.2	6.8	2266	15	US-10-108-260A-375	Sequence 375, App	
C	23	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App	
C	22	31.4	6.9	7374	14	US-10-505-428-1006	Sequence 10067, A	
C	21	31.4	6.9	7374	10	US-09-764-891-10207	Sequence 10207, A	
C	20	31.4	6.9	460	15	US-10-260-238-787	Sequence 787, App	
C	19	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, App1	
C	18	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A	
C	17	31.8	7.0	930	12	US-10-382-122A-20335	Sequence 20335, A	
C	16	31.8	7.0	930	15	US-10-382-122A-20430	Sequence 20430, A	
C	15	31.8	7.0	930	12	US-10-369-493-24480	Sequence 24480, A	
C	14	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A	
C	13	31.6	7.0	978	15	US-10-369-493-24265	Sequence 24265, A	
C	12	31.4	6.9	460	15	US-10-260-238-787	Sequence 787, App	
C	11	31.4	6.9	7374	10	US-09-764-891-10207	Sequence 10207, A	
C	10	31.4	6.9	7374	14	US-10-505-428-1006	Sequence 10067, A	
C	9	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App	
C	8	31.2						

ALIGNMENTS

```

RESULT 1
US-10-156-761-3483/c
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMODA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 3483
LENGTH: 1662
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1662)
US-10-156-761-3483

Query Match          7.6%; Score 34.6; DB 14; Length 1662;
Best Local Similarity 49.7%; Pred. No. 0.78;
Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0

Cy 50 GTGCTTCAGCGGGGCTGCTTCACACATGAGGGGCGGGCGGGTATCATATACGCGCGGCGCA 109
Db 590 GCGGCGGGGTGAGCTCGATACAGATCGCGACGCCGTGTTTCATCTGTCACGCGGAGCGGC 531

```

QY 110 ATAGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTCAACGCTGGCC 169
 DB 530 TGGGGGCGCCGACATGCGCGGAGTCCGGCGGTGATCTGCGCGGAGGGTGGC 471
 QY 170 TTGCTTCGCAAGCGATGCGCGTAATATGATCAGTTCGGTTCGCTGTTTACCC 226
 DB 470 TTGAACCTTCTGGGGGCGAGCGCGGAGACGCTCTCGTACGTCCTCGAGGATGCC 414

RESULT 2

US-10-156-761-1
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermectilis
 ; FEATURE:
 ; NAME/KEY: misc-feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Query Match
 Best Local Similarity 7.6%; Score 34.6; DB 14; Length 9025608;
 Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 50 GTGCTGTGCTGCGCTGCTTCACATGAGGGCGGGGTAATCATACGCGCGGCA 109
 DB 4331781 GCGCGCGGGTCCGACGTCGATACAGATCGGACGCCGCTTCATCGTACGCGCGGCGCGC 4331840
 QY 110 ATAGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTTCGCTAAGCTGCGC 169
 DB 4331841 TGGGGGCGCCGACATGCGCGGAGTCCGGCGGTGAGGGTATCGTCCGGGAGGGTGGC 4331900
 QY 170 TTGCTTCGCAAGCGATGCGCGTAATATGATCAGTTCGGTTCGCTGTTTACCC 226
 DB 4331901 TTGAACCTTCTGGGGGCGAGCGCGGAGACGCTCTCGTACGTCCTCGAGGATGCC 4331957

RESULT 3

US-09-835-684-6/c
 ; Sequence 6, Application US/09835684
 ; Patent No. US20020019337A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Qiu, Dewen
 ; APPLICANT: Remick, Dean
 ; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
 ; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
 ; FILE REFERENCE: 21829/71
 ; CURRENT APPLICATION NUMBER: US/09/835,684
 ; PRIOR FILING DATE: 2001-04-16
 ; PRIOR APPLICATION NUMBER: 60/198,359
 ; PRIOR FILING DATE: 2000-04-19

NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora
 US-09-835-684-6

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGCTGCGCTGCTTCACATGAGGGCGGGGTAATATACGCGCGCAAT 111
 DB 209 GCTCGGTTGGCCGATTAACCTGATTTGGCGGATAGAGTACTTAAAGTTCCGCAAT 150
 QY 112 AGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTTCGCTAAGCTGCGCTT 171
 DB 149 AATTGACCATTTGCTCAATGATGTTGCCGATGATGATGTTGCCCCAACCGCAATTT 90
 QY 172 GC 173
 DB 89 GC 88

RESULT 4

US-09-880-371-6/c
 ; Sequence 6, Application US/09880371
 ; Patent No. US20020059658A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Derocher, Jay
 ; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
 ; FILE REFERENCE: 21829/91
 ; CURRENT APPLICATION NUMBER: US/09/880,371
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: 60/211,585
 ; PRIOR FILING DATE: 2000-06-15
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora
 US-09-880-371-6

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGCTGCGCTGCTTCACATGAGGGCGGGGTAATATACGCGCGCAAT 111
 DB 209 GCTCGGTTGGCCGATTAACCTGATTTGGCGGATAGAGTACTTAAAGTTCCGCAAT 150
 QY 112 AGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTTCGCTAAGCTGCGCTT 171
 DB 149 AATTGACCATTTGCTCAATGATGTTGCCGATGATGATGTTGCCCCAACCGCAATTT 90
 QY 172 GC 173
 DB 89 GC 88

RESULT 5

US-09-879-248-5/c
 ; Sequence 5, Application US/09879248
 ; Patent No. US20020062500A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fan, Hao
 ; APPLICANT: Wei, Zhong-Min
 ; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
 ; TITLE OF INVENTION: THEREOF

Qy 18 GCGAGCATTCGACAGCATCGTAGTTTCTGCAGTGCTCTGGCTGGGCGTCGTCCACCAATC
::: ||| : : :::: :

```

1 Sequence 418, Application US/10184634
2 Publication No. US20030066841
3 GENERAL INFORMATION:
4 APPLICANT: Baker, Kevin P.
5 APPLICANT: Chen, Dian
6 APPLICANT: Desnoyers, Luc
7 APPLICANT: Goddard, Audrey
8 APPLICANT: Godowski, Paul J.
9 APPLICANT: Gurney, Austin L.
10 APPLICANT: Pan, James
11 APPLICANT: Smith, Victoria
12 APPLICANT: Watanabe, Colin K.
13 APPLICANT: Wood, William I.
14 APPLICANT: Zhang, Zemin
15 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
16 FILE REFERENCE: P3430R1C217
17 CURRENT FILING DATE: 2002-06-28
18 NUMBER OF SEQ ID NOS: 612
19 SEQ ID NO 418
20 LENGTH: 655
21 TYPE: PRT
22 ORGANISM: Homo Sapien
23 IS-10-184-634-418

```

Best Local Similarity .53; score 33.4; DB 14; Length 655;
Matches 39; Contamination 14.9%; Pred. No. 1.2;

18 GGCAGCAATCCGACCAATTTT
 100% Mismatches 147; Indels 0; Gaps 0;

70 000000...SS.B 26

DYDD.C..DNBB.N.AMNBH.MCYTSSANH.BAC...SS.B 26

CSM.COAI.KS...M.

265 M.YBSTSSNS.ASNKCM.BRASS.B.DA.KD.Y.TAC.S.SASSS.S.ACGBNHA 206

205 MAC.M.BB.N.S.CHARTMKNKISS.BHS.ACYTG.MGYBH.A.ACMWYBCTCH.CQAT

[illegible]

85 .SANAGD.SBRAC T GMHGA P 64

100-368601-204

Sequence 7222, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT:

APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HITROSHI

APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOISE-RE-

CURRENT AFFILIATION NUMBER: U
CURRENT FILING DATE: 2002-05
PRIOR AFFILIATION:

PRIOR APPLICATION NUMBER: JP
PRIOR FILING DATE: 2001-08-02

LENGTH: 1848
TYPE: DNA

NAME/KEY: CDS
LOCATION: 4A

Query Match 7.38:

46

105 CCGGAGGATGTTT

1/4 GTGACGGTTTCCGGCGGGGA

714 GGGGTGCGCACAGGTGG

654 GCCGAGCGGCGCAGCAGGT

10-156-761-1/c

GENERAL INFORMATION:
APPLICANT: OMTIBA SATOSHI

APPLICANT: HORIKAWA, HIROSHI
APPLICANT: HORIKAWA, JUN

INVENTOR: HATTORI, MASAHIRA
ATTORNEY: NOVEL POLYNTIC

CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: TD 0204

REGISTRATION NUMBER: JP 2001
PRIOR FILING DATE: 2001-08-02
TIMBER OF SEQ ID NO:

PRIOR FILING DATE: 2001-08-02
 INVENTOR NAME: JY 2001-272697
 NUMBER OF SEQ ID NOS: 15109
 ID NO 1


```

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6028
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(930)
US-09-815-242-6028

```

```

Query Match          7.0%; Score 31.8; DB 9; Length 930;
Best Local Similarity 52.7%; Pred. No. 5.2;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 311 ACCAGTGAACCTTAAACTCCGATATTAATGCTGCGCAATACGCGGTATTAACGCCG 370
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233 AGCAGCAGCGCGGTAAAGACCGCAATGATGCGCGGTACAGCCTGGGGGAATACCTCG 292
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 371 CGCTGTATATCAGACCGCATCTGATTCACGCTAATGCTGCTGAGTTGTTGGCA 430
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 293 CGCTGTGTTGGCGCTGTGATTTGATTTGCTGATGCGGTGCTGTGAGATGCGCG 352
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 ACAAGCCACG 441
    ||||| |||||
DB 353 GCAAGTTCATG 363
    ||||| |||||

```

Search completed: March 17, 2004, 08:16:43
 Job time : 424.736 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds

(without alignments) 6103.863 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaacttttaaaagfsgc.....ccacgctaaccagctactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqe, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

```

EST:
1: em_estba:*
2: em_estin:*
3: em_estin:*
4: em_estin:*
5: em_estin:*
6: em_estin:*
7: em_estin:*
8: em_estin:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est6:*
15: gb_est7:*
16: gb_est8:*
17: gb_est9:*
18: gb_est10:*
19: gb_est11:*
20: gb_est12:*
21: gb_est13:*
22: gb_est14:*
23: gb_est15:*
24: gb_est16:*
25: gb_est17:*
26: gb_est18:*
27: gb_est19:*
28: gb_est20:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query length	ID	Description
C 1	237.4	52.1	680	12 BU618688
C 2	37.4	8.2	484	9 AV430994
C 3	35.4	7.8	664	13 CA141136
C 4	34.8	7.6	675	13 BX251172

Result No.	Score	Query length	ID	Description
C 5	34.6	7.6	699	12 B1959328
C 6	34.2	7.5	296	9 AV058630
C 7	34.2	7.5	669	28 A2959284
C 8	34.2	7.5	978	28 B2569281
C 9	34.2	7.5	1036	12 BG671547
C 10	34.2	7.5	1137	28 B2565783
C 11	34.2	7.5	234	10 B8581987
C 12	34.2	7.5	500	14 CA711477
C 13	34.2	7.5	500	14 CA711477
C 14	33.8	7.4	365	9 AA746477
C 15	33.8	7.4	559	14 CD374421
C 16	33.8	7.4	559	13 BU765156
C 17	33.8	7.4	902	28 B2569250
C 18	33.8	7.4	941	29 CG178280
C 19	33.8	7.4	1098	29 CNG02GDP
C 20	33.6	7.4	680	10 BF668888
C 21	33.6	7.4	966	13 BQ718166
C 22	33.4	7.3	505	6 AL809250
C 23	33.4	7.3	523	12 BJ334624
C 24	33.4	7.3	715	13 BU444683
C 25	33.4	7.3	1064	13 BQ939900
C 26	33.2	7.3	317	12 BM333324
C 27	33.2	7.3	625	10 BF501499
C 28	33.2	7.3	629	9 AU295325
C 29	33.2	7.3	806	13 BX888120
C 30	33.2	7.2	509	12 BM896625
C 31	33.2	7.2	649	12 B1261742
C 32	33.2	7.2	826	28 B2462870
C 33	32.8	7.2	521	14 CA703231
C 34	32.8	7.2	572	12 B1626794
C 35	32.8	7.2	594	12 BJ335653
C 36	32.8	7.2	613	14 CB935126
C 37	32.8	7.2	653	12 BU336964
C 38	32.8	7.2	1893	28 AY414592
C 39	32.6	7.1	204	28 A2577991
C 40	32.6	7.1	352	9 AA784257
C 41	32.6	7.1	356	9 AA785291
C 42	32.6	7.1	778	29 CNG03735
C 43	32.6	7.1	813	28 B2204853
C 44	32.6	7.1	919	14 CK246169
C 45	32.6	7.1	942	14 CK247920

ALIGNMENTS

RESULT 1
LOCUS BU618688/680 bp mRNA linear EST 01-OCT-2003
DEFINITION BU618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone X1186D22 5', mRNA sequence.
ACCESSION BU618688
VERSION BU618688.1 GI:37256713
KEYWORDS EST.

SOURCE Xenopus laevis (African clawed frog)

ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 680)
Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin, I., T. and Kohara, Y.
Expressed genes in X. laevis embryo

TITLE JOURNAL
COMMENT Unpublished (2001)
Contact: Tadao Shin-1
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6886
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.

FEATURES
source

http://xenopus.nibb.ac.jp.
Location/Qualifiers
1. 680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="X1186D22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 52.1%; Score 237.4; DB 12; Length 680;
Best Local Similarity 70.0%; Pred. No. 6,4e-57;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTCGACGACATGATGTTCTGGACGAGTCTGCT 60
DB 593 ATGAACCTTTAAAGTGGAGCATTCGACGACATGATGTTCTGGACGAGTCTGCT 60
QY 61 GGGCTCTTCCAAATAGGGGCGCGCGGTATCATACGGCGCGCATATAGTCCGCG 120
DB 533 GGGTGTCTTCTCAGTACGCGCGCGGTATCATACGGCGCGCATATAGTCCGCG 120
QY 121 CCGGACTCAACGTGAGCATTTATCATGTCGTTCCGCTAACGCTGCGCTGCTGCA 180
DB 473 CCAATTTGAGTGAACATTTACAGTACGTCGCGGTAACTTTCGCTGCA 414
QY 181 AGCGATCCCGTAAATATGATCATGTCGTCGCTGTTGTTACCATGAATGCAAT 240
DB 413 ACTGATGCCCCGTAACCTGACTGATCACTATTAACGAGATGCGCGGTATGTCGAT 354
QY 241 GAGGCGCAGGTGCGGATATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 353 GTTGGTCAGGGCTCAGATGACGCTCATGATGATGATGATGATGATGATGATGAT 294
QY 301 GCCACCATCCGACGATGAAAGCTAAAGCTCCGATATGATGATGATGATGATGATGAT 360
DB 293 GCTACTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234
QY 361 AATAACCGCGCGGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 233 GGCACCGGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 174
QY 421 GGTTCGCAACGACCGCATGATGATGATGATGATGATGATGATGATGATGATGAT 456
DB 173 GCGTTGTAACAGCGGACCGCTCATGATGATGATGATGATGATGATGATGATGAT 138

RESULT 2
AV430994/c

LOCUS 484 bp mRNA linear EST 23-AUG-2000
DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
TITLE 1 (bases 1 to 484)
AUTHORS Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
Tabata, S.
JOURNAL Generation of 10,154 expressed sequence tags from a leafy
MEDLINE Gematophyte of a marine red alga, Porphyra yezoensis
PUBMED 20363100
10907854
CONTACT: Erika Asamizu
COMMENT The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute

FEATURES
source

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. 484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="PI027e10.r"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptSK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 10;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 321 CGTAAACTCCGATATTAATGTCGCGCAATACGCGGTATTAACCGCGCTGTTAA 380
DB 441 CCAAAACACACAGTGTCCGCGCTCAACGCGCTCAATACGCGCTGTCGCGAG 382
QY 381 TCAGCCGATCGATTCAGCGTATGTCGTCAGTGTGTTTGGCAACACGCA 439
DB 381 GCCGAAGCTGTGTATCCAAAGTATGTCGTCAGGTCGCGGAAGGCGCTCCCGCA 323

RESULT 3

CA141136/c
LOCUS 664 bp mRNA linear EST 24-SEP-2003
DEFINITION SCURFT058D10.9 RT2 Saccharum officinarum cDNA clone SCURFT058D10
ACCESSION CA141136
VERSION CA141136.1 GI:35034821
KEYWORDS EST
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
1 (bases 1 to 664)
Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.
The libraries that made SUCST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parnuda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.iteav.unesp.br
Plate: 058 row: D column: 10
Seq primer: T7 Promoter Primer.
Location/Qualifiers
1. 664
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCURFT058D10"
/lab_host="DH10B"
/clone_lib="RT2"
/note="Organ: Root tips (0.3cm-long) from adult plants;
vector: pSPori; Site_1: SalI; Site_2: NotI; An
unidirectional cDNA library generated from (Root
from polyA+ mRNA using Superscript Plasmid System kit
(Invitrogen). The double-strand cDNAs were fractionated
in a Sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 kb were directionally cloned into the
vector. Details of each source of RNA and library

construction can be obtained at
http://suceest.lad.ic.unictamp.br/public"

Query Match 7.8%; Score 35.4; DB 13; Length 664;
Best Local Similarity 46.2%; Pred. No. 44;
Matches 117; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 22 GCATTGGAGCAAAATGTAAGTTTCTGGCAGTGTCTGGCTGGGCTGTCCCAATAGGGGC 81
DB 521 GCTTCGAGTATGTCGAGCTGAGCACCCTGTTGCGTTCAGACCTTGTACATGCGCAC 462
QY 82 GCGCGCGGTAATCATATACGCGCGCGCAATAGTTCCGCGCGCACTCAACGTTAGCATT 141
DB 461 GTGCGCGGCGGCGCGCATCCGCGCGCGACCCCGCGCTGGCCCAAAACCTGAGCCCC 402
QY 142 TATCAGTACGGTTCCGCTTACGCTGCGCTTCTGCAAGAGATGCCCGTAATATGAT 201
DB 401 CTGACGACGCGCTCGGTCGCGCTGCTGCGGCTGCTGCGCGCTGCGCGCTCATGCAC 342
QY 202 CAGCTGCTTACCCGCTGTTTATACCATGAAATGACATGACGCGCAGGTCGATAT 261
DB 341 CGGCGACGAGAGTTGCTGCGCGCGGTGGGACGAGCATGAAGAGCGAGCTCCGATGAG 282
QY 262 AGTACTATTGAAC 274
DB 281 CTGTGTTGCAC 269

RESULT 4 675 bp mRNA linear EST 24-FEB-2003
BX251172
LOCUS
DEFINITION BX251172 Pinus pinaster differentiating xylem adult Pinus pinaster
CDNA clone P0046C03, mRNA sequence.
ACCESSION BX251172
VERSION BX251172.1 GI:28513831
KEYWORDS EST.
SOURCE Pinus pinaster
ORGANISM Pinus pinaster
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
REFERENCE 1 (bases 1 to 675)
Canton, F.R., Le Provost, G., Garcia, V., Barre, A., Frigerio, J.-M.,
Dauvergne, A., Canovas, F.M., and Plomion, C.
Transcriptome analysis of wood formation in maritime pine
Unpublished (2003)
TITLE Contact: Frigerio JM
JOURNAL Genetic et Amelioration 69
COMMENT INRA

route d'Arcachon 33612 Cestas CEDEX France
Email: Frigerio@pierroton.inra.fr
Email: Frigerio@pierroton.inra.fr
Location/Qualifiers

FEATURES
source
1..675
/organism="Pinus pinaster"
/mol_type="mRNA"
/strain="ecotype: Corsican"
/db_xref="taxon:71647"
/clone="P0046C03"
/tissue_type="differentiating xylem"
/dev_stages="adult"
/clone_lib="Pinus pinaster differentiating xylem adult"
/note="Vector: Uni-zap XR lambda (Stratagene); Site 1: Eco
RI; Site 2: Xho I; A composite cDNA library was made with
mRNA isolated from normal, compression, opposite, early
and late wood of Maritime pine uni-directionally cloned
into Uni-zap XR using the ZAP-CDNA Synthesis kit
(Stratagene). pBluescript SK(-) plasmids were obtained by
in vivo mass excision. The nucleotide sequence of the
5' end was obtained by automated sequencing with the T3
primer by GENOME EXPRESS, Meylan, France"

ORIGIN

Query Match 7.6%; Score 34.8; DB 13; Length 675;
Best Local Similarity 57.3%; Pred. No. 66;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 192 TAAATATGATGAGGTGTTACCGTGTGTTACCATTAATGACATGACGAGCCAGG 251
DB 526 TAAATTCCTAGATAGTATGTTATTTTATTTATGACATGACATGACATGCTGCTGG 585
QY 252 TGGGATTAATGACTATTTGAACTGACTGAGATGTTTGAATAATATG 301
DB 586 TTTCGATGATGTAGAGATGTAACAACATATACCCGACATGATATATG 635

RESULT 5 699 bp mRNA linear EST 22-OCT-2001
BX1959328/c
LOCUS
DEFINITION BX1959328 Hordeum vulgare rachis EST library HVCDNA0015
(normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEN0019E09f,
mRNA sequence.
ACCESSION BX1959328
VERSION BX1959328.1 GI:16310583
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 699)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,
Simmons, J., Oates, R., and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex rachis cDNA library
Unpublished (2001)
CONTACT: Wing RA
JOURNAL Clemson University Genomics Institute
COMMENT 100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total bp bases = 275
Seq primer: AATTAACCTCCTACATTAACG
High quality sequence stop: 452.
Location/Qualifiers

FEATURES
source

1..699
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEN0019E09f"
/tissue_type="Rachis"
/lab_host="TJ121"
/clone_lib="Hordeum vulgare rachis EST library HVCDNA0015
(normal)"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Plants were grown at Washington State University,
Pullman, WA in a greenhouse, the rachises were excised and
frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close
lab at the University of California, Riverside total RNA
was prepared, poly(A) was purified, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The
sequence has been trimmed to remove vector sequence and
contains a minimum of 100 bases of piped value 20 or
above. For more details on library preparation and

```

location/qualities
1. 296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="181:0054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"

```

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="DUGC2M0226G20"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_1ib="Mouse 10kb plasmid DUGC2M library"
/notes="Vector: PMD219; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adsorbed DNA was purified and size-selected for a 9.5 to

```


RESULT 11	
BB581987/c	
LOCUS	214 bp
DEFINITION	BB581987 RIKEN full-length enriched, adult male colon Mus musculus EST 30-NOV-2006
ACCESSION	CDNA clone J030005005.5, mRNA sequence.
VERSION	BB581987
KEYWORDS	BB581987.1 GI:11478531
SOURCE	EST.
	Mus musculus (house mouse)

Query Match 7.5%; Score 34; DB 10; Length 234

1..234
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="taxon:10090"
/clone="9030005005"
/sex="male"
/tissue_type="colon"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAATTCACAGCCCTTTTTTTTTTTTNN 3'], cDNA was prepared by using rRNAase thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAATTCGCAGTGTAATTAAATATCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I. Cloning sites, 5' end: SalI; 3' end: BamHI"

Beet Local Similarity 51.3%; Pred. No. 70;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 35 TCGTAGTTTGGCAGTGTCTGCTGGCGCTTCCACATAGGGCGCGGTAATC 94
Db 200 TAGTTTGGTAGGTGCGTTATGCGGGTTCCTTCACACAGCCCGCGCTTCC 141

Qy 95 ATACGGGGGGGCAATAGTTCGGCGCGGACTCAAGCTTATGAGTATGAGTGT 154
Db 140 CGAACCGCGGCTTTATAGCTTGTTCMAAATGAGGAGCATTAACCACTT 81

Qy 155 CCGCTACGCTGCGCTTGTCTGCAAGCATGC 188
Db 80 CCGGTACCAAGCCGCGCGCATTCGCGGC 47

RESULT 12
CA711477 500 bp mRNA linear EST 26-NOV-2002
LOCUS wdk2c.pk014.03 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.03 5'
DEFINITION end, mRNA sequence.
ACCESSION CA711477.1 GI:25433270
VERSION CA711477.1 GI:25433270
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 500)
Tingley,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE Dupont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingley
Crop Genetics
E. I. Dupont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingley@USA.dupont.com
Seq primer: M13.

FEATURES
source 1..500
Location/Qualifiers
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wdk2c.pk014.03"
/tissue_type="kernel"
/clone_lib="wdk2c"
/note="Vector: pBluescript SK+, Site_1: EcoRI; Site_2:
XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
days after anthesis."

ORIGIN
Query Match 7.5%; Score 34; DB 14; Length 500;
Best Local Similarity 54.5%; Pred. No. 97;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 15 AGTGCACATTCGAGCAATCGTAGTTCTGCGAGTCTGCTGCGCGTTCACACA 74
Db 274 AGTGNATGACATCGGAGCGCNCGGTCTCCCGCGCGTTCGAAACCGCGGAGCTCCCN 215

Qy 75 ATGGGGCGCGCGGTATATCATACGCGCGCGCAATAGTTCGCGCGCGGAC 126
Db 214 TGTGGAGCGGTGGCATGACATGCGCGCGCGCGCGGTGGCGCCAC 163

RESULT 13
LOCUS A2180696 500 bp DNA linear GSS 30-AUG-2000
DEFINITION SP_0181_A1_c06 Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus

ACCESSION A2180696
VERSION A2180696.1 GI:8353071
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus
ORGANISM Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoida; Euechinoida; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 500)
Cameron,R.A., Mahatras,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Pousetka,A.D., Livingston,B.T.,
Wray,G.A., Betensohn,C.A., Lehrach,H., Britten,R.J., Davidson,E.H.
and Hood,L.
TITLE A sea urchin genome project: Sequence scan, virtual map, and
additional resources
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
PUBMED 10920195
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-1047
Email: acameron@caltech.edu
Plate: 181 row: E column: 11
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 500.

FEATURES
source 1..500
Location/Qualifiers
/organism="Strongylocentrotus purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="plate=181 Col=11 Row=E"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC6.6; BAC clones in E-Coli
DH10B"

ORIGIN
Query Match 7.5%; Score 34; DB 28; Length 500;
Best Local Similarity 50.3%; Pred. No. 97;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 133 TTGAGCATTTATCAGTACGCTTCGCTGCAAGGATCCCGT 192
Db 8 TTNTCCATATATATCCATGATTAACGATGATGATGATGATGATGATGA 67

Qy 193 AAATATGATCAGCTGTTACCCGCTGTTTACCATGAATGCAATGACGCGCAGGT 252
Db 68 GACCATGATGCGATGATGATGATGATGATGATGATGATGATGATGAT 127

Qy 253 GCGATATAGTACTATTGAACTGACTCAGAAATGTTTCGAA 295
Db 128 GATATGATATATATATGAAATTAATTAATGATGATGATGATGATGAT 170

RESULT 14
LOCUS AA746477 365 bp mRNA linear EST 27-JAN-1998
DEFINITION hw62e02.g1 NCI CGAP GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
similar to TR:000555 000555 NEURONAL CALCIUM CHANNEL ALPHA 1A
SUBUNIT ISOFORM A-1; contains element TARI repetitive element ;
mRNA sequence.
ACCESSION AA746477
VERSION AA746477.1 GI:2786463
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using BW model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgsc.....ccacgcctaacagttatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA Main:*

```

1: /cgn2_6/prodata/2/pna/US06_PCTUS_COMB.seq:*
2: /cgn2_6/prodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/prodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/prodata/2/pna/US08_COMB.seq:*
5: /cgn2_6/prodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/prodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/prodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/prodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/prodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/prodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/prodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/prodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/prodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/prodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/prodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/prodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/prodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/prodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/prodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/prodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/prodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/prodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/prodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/prodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/prodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/prodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/prodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/prodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/prodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/prodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/prodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/prodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/prodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/prodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/prodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/prodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/prodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/prodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/prodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/prodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/prodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/prodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/prodata/2/pna/US100A_COMB.seq:*

```

```

44: /cgn2_6/prodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/prodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/prodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/prodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/prodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/prodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/prodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/prodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/prodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/prodata/2/pna/US105_COMB.seq:*
54: /cgn2_6/prodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/prodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/prodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/prodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/prodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/prodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/prodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/prodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/prodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/prodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/prodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/prodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/prodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/prodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/prodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/prodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/prodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/prodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/prodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/prodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/prodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/prodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/prodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/prodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/prodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/prodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/prodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/prodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/prodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/prodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/prodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/prodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/prodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/prodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/prodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/prodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/prodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/prodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/prodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/prodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/prodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/prodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/prodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/prodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/prodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/prodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/prodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/prodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/prodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/prodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/prodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/prodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/prodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/prodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/prodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/prodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/prodata/2/pna/US6053_COMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543.407
;; CURRENT FILING DATE: 2000-04-05
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 1
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: *Salmonella enteritidis*
US-09-543-407-1

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.9e-115;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTTCTGGAGTGTCTGCT 60
Db 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTTCTGGAGTGTCTGCT 60
Qy 61 GCGCTGTTCCACAATGGGGCGGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCACAATGGGGCGGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGCTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTTGC 180
Db 121 CCGGACTCAAGCTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTTGC 180
Qy 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTACGGCGCGAT 240
Db 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTACGGCGCGAT 240
Qy 241 GTAGCGCAGGGTGGCATTAATAGTACTTGAATCTGACTCAGATGATTTTCAAAATAT 300
Db 241 GTAGCGCAGGGTGGCATTAATAGTACTTGAATCTGACTCAGATGATTTTCAAAATAT 300
Qy 301 GCCACATCGACCAATGGAAAGCTTAAATCTCCGATATTAATGTCGGCAATACGGCGGT 360
Db 301 GCCACATCGACCAATGGAAAGCTTAAATCTCCGATATTAATGTCGGCAATACGGCGGT 360
Qy 361 AATAAGCGCGGCTGTTAATCAAGCCGATCTGATTCAGAGTAAAGTGTGCGTCA 420
Db 361 AATAAGCGCGGCTGTTAATCAAGCCGATCTGATTCAGAGTAAAGTGTGCGTCA 420
Qy 421 GGTITGGCAACAACGCGCGCTAACCAGATTTAA 456
Db 421 GGTITGGCAACAACGCGCGCTAACCAGATTTAA 456

RESULT 3
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of *Leishmania major*.

US-09-543-407-19

Query Match 90.5%; Score 412.8; DB 23; Length 456;
Best Local Similarity 94.1%; Pred. No. 2.2e-121;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTTCTGGAGTGTCTGCT 60
Db 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTTCTGGAGTGTCTGCT 60
Qy 61 GCGCTGTTCCACAATGGGGCGGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCACAATGGGGCGGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGCTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTTGC 180
Db 121 CCGGACTCAAGCTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTTGC 180
Qy 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTACGGCGCGAT 240
Db 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTACGGCGCGAT 240
Qy 241 GTAGCGCAGGGTGGCATTAATAGTACTTGAATCTGACTCAGATGATTTTCAAAATAT 300
Db 241 GTAGCGCAGGGTGGCATTAATAGTACTTGAATCTGACTCAGATGATTTTCAAAATAT 300
Qy 301 GCCACATCGACCAATGGAAAGCTTAAATCTCCGATATTAATGTCGGCAATACGGCGGT 360
Db 301 GCCACATCGACCAATGGAAAGCTTAAATCTCCGATATTAATGTCGGCAATACGGCGGT 360
Qy 361 AATAAGCGCGGCTGTTAATCAAGCCGATCTGATTCAGAGTAAAGTGTGCGTCA 420
Db 361 AATAAGCGCGGCTGTTAATCAAGCCGATCTGATTCAGAGTAAAGTGTGCGTCA 420
Qy 421 GGTITGGCAACAACGCGCGCTAACCAGATTTAA 456
Db 421 GGTITGGCAACAACGCGCGCTAACCAGATTTAA 456

RESULT 4
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from Gp63 of *Leishmania major*.
US-09-543-407-17

Query Match 89.1%; Score 406.4; DB 23; Length 456;
Best Local Similarity 93.2%; Pred. No. 2.5e-119;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTTCTGGAGTGTCTGCT 60
Db 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTTCTGGAGTGTCTGCT 60
Qy 61 GCGCTGTTCCACAATGGGGCGGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120

```
Db 61 GGCCTCGTTCCAAAGGCGCGCGCGGTAATACGCGCGCGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGTAACGCTGCTTGCGCA 180
Db 121 CCGGACTATGATCAGCTGGTTACCGGTGTTGTTACCCATGTAATATGACATGCACTGCA 180
Qy 181 AGCGATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTTACGCGCGCGAT 240
Db 181 AGCGATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTTACGCGCGCGAT 240
Qy 241 GTAGGCGAGGGTGGCGATTAATAGTACTTAACTGACTGAAATGTTTCAAGAAATAT 300
Db 241 GTAGGCGAGGGTGGCGATTAATAGTACTTAACTGACTGAAATGTTTCAAGAAATAT 300
Qy 301 GCCACCATGACCAAGTGAACGCTAAATCCGATTAATCTGTGCGGCAATAGCGCGGT 360
Db 301 GCCACCATGACCAAGTGAACGCTAAATCCGATTAATCTGTGCGGCAATAGCGCGGT 360
Qy 361 AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGTTAATGTCGTCAGGTT 420
Db 361 AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGTTAATGTCGTCAGGTT 420
Qy 421 GGTGTTGGCAACCAACCGCGCTAACCGATTTAA 456
Db 421 GGTGTTGGCAACCAACCGCGCTAACCGATTTAA 456
```

```
RESULT 5
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-11
```

```
Query Match      88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
Qy 1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTGCGCTGCTGCT 60
Db 1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTGCGCTGCTGCT 60
Qy 61 GCGCTGTTCCAAATGCGGCGCGCGGTATCATTAACGCGCGCGCAATAGTTCCGCG 120
Db 61 GCGCTGTTCCAAATGCGGCGCGCGGTATCATTAACGCGCGCGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGTAACGCTGCTTGCGCA 180
Db 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGTAACGCTGCTTGCGCA 180
Qy 181 AGCGATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTTACGCGCGCGAT 240
Db 181 AGCGATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTTACGCGCGCGAT 240
```

```
Qy 241 GTAGGCGAGGGTGGCGATTAATAGTACTTAACTGACTGAAATGTTTCAAGAAATAT 300
Db 241 GTAGGCGAGGGTGGCGATTAATAGTACTTAACTGACTGAAATGTTTCAAGAAATAT 300
Qy 301 GCCACCATGACCAATGGAACGCTAAATCCGATTAATCTGTGCGGCAATAGCGCGGT 360
Db 301 GCCACCATGACCAATGGAACGCTAAATCCGATTAATCTGTGCGGCAATAGCGCGGT 360
Qy 361 AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Db 361 AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Qy 421 GGTGTTGGCAACCAACCGCGCTAACCGATTTAA 456
Db 421 GGTGTTGGCAACCAACCGCGCTAACCGATTTAA 456
```

```
RESULT 6
US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-13
```

```
Query Match      88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
Qy 1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTGCGCTGCTGCT 60
Db 1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTGCGCTGCTGCT 60
Qy 61 GCGCTGTTCCAAATGCGGCGCGCGGTATCATTAACGCGCGCGCAATAGTTCCGCG 120
Db 61 GCGCTGTTCCAAATGCGGCGCGCGGTATCATTAACGCGCGCGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGTAACGCTGCTTGCGCA 180
Db 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGTAACGCTGCTTGCGCA 180
Qy 181 AGCGATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTTACGCGCGCGAT 240
Db 181 AGCGATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTTACGCGCGCGAT 240
Qy 241 GTAGGCGAGGGTGGCGATTAATAGTACTTAACTGACTGAAATGTTTCAAGAAATAT 300
Db 241 GTAGGCGAGGGTGGCGATTAATAGTACTTAACTGACTGAAATGTTTCAAGAAATAT 300
Qy 301 GCCACCATGACCAAGTGAACGCTAAATCCGATTAATCTGTGCGGCAATAGCGCGGT 360
Db 301 GCCACCATGACCAAGTGAACGCTAAATCCGATTAATCTGTGCGGCAATAGCGCGGT 360
Qy 361 AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Db 361 AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
```


Qy 421 GATTGGCAACAACGCGCTAACGATTTAA 456
Db 421 GATTGGCAACAACGCGCTAACGATTTAA 456

RESULT 7

US-09-543-407-23
Sequence 23, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-23

Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
Qy 61 GGCCTGTTCCACATGCGGCGCGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
Db 61 GGCCTGTTCCACATGCGGCGCGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
Qy 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTACGCTGCTCTGCAA 180
Db 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTACGCTGCTCTGCAA 180
Qy 181 AGCGATCGCCGTAATCTGAACGACATTCACGAGCGGTTATGTAGCGGCGCAT 240
Db 181 AGCGATCGCCGTAATCTGAACGACATTCACGAGCGGTTATGTAGCGGCGCAT 240
Qy 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAAGTGAAGTGGTTGAGAAATAT 300
Db 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAAGTGAAGTGGTTGAGAAATAT 300
Qy 301 GCCACCATCGACGCTGGAAGCGTAAATCTCCGATATTAATGTCGGCAATAGCGCGT 360
Db 301 GCCACCATCGACGCTGGAAGCGTAAATCTCCGATATTAATGTCGGCAATAGCGCGT 360
Qy 361 AATAAGCGCGGCGTGAATCAGACCGCATCTGATTCAGAGTAAATGTTGAGTT 420
Db 361 AATAAGCGCGGCGTGAATCAGACCGCATCTGATTCAGAGTAAATGTTGAGTT 420
Qy 421 GATTGGCAACAACGCGCTAACGATTTAA 456
Db 421 GATTGGCAACAACGCGCTAACGATTTAA 456

RESULT 8
US-09-543-407-25
Sequence 25, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-25

Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
Qy 61 GGCCTGTTCCACATGCGGCGCGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
Db 61 GGCCTGTTCCACATGCGGCGCGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
Qy 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTACGCTGCTCTGCAA 180
Db 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTACGCTGCTCTGCAA 180
Qy 181 AGCGATCGCCGTAATCTGAACGACATTCACGAGCGGTTATGTAGCGGCGCAT 240
Db 181 AGCGATCGCCGTAATCTGAACGACATTCACGAGCGGTTATGTAGCGGCGCAT 240
Qy 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAAGTGAAGTGGTTGAGAAATAT 300
Db 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAAGTGAAGTGGTTGAGAAATAT 300
Qy 301 GCCACCATCGACGCTGGAAGCGTAAATCTCCGATATTAATGTCGGCAATAGCGCGT 360
Db 301 GCCACCATCGACGCTGGAAGCGTAAATCTCCGATATTAATGTCGGCAATAGCGCGT 360
Qy 361 AATAAGCGCGGCGTGAATCAGACCGCATCTGATTCAGAGTAAATGTTGAGTT 420
Db 361 AATAAGCGCGGCGTGAATCAGACCGCATCTGATTCAGAGTAAATGTTGAGTT 420
Qy 421 GATTGGCAACAACGCGCTAACGATTTAA 456
Db 421 GATTGGCAACAACGCGCTAACGATTTAA 456

RESULT 9
US-09-543-407-15
Sequence 15, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gpe3 of *Leishmania* major.
US-09-543-407-15

Query Match 88.4%; Score 403.2; DB 23; Length 456;
Best Local Similarity 92.5%; Pred. No. 2,7e-118;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCGCAATGCTGTTCTGCGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGCGCAATGCTGTTCTGCGCTGCTGCT 60
QY 61 GGGGTCTTCCACATATGGGCGGCGGCTGATCATACGGGCGGCAATGTTCCGGC 120
DB 61 GGGGTCTATGATCAGCTGGTTACCGGTGTTGTTACCATGAATGCGCATGCTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 AGCGATGCCGCTGAAATCTGAAAAGCATTAACCGAGGCGGTTATGTTAAAGGCGCGAT 240
DB 181 AGCGATGCCGCTGAAATCTGAAAAGCATTAACCGAGGCGGTTATGTTAAAGGCGCGAT 240
QY 241 GTAGGCGGAGGTGGGATTAATGATCTATTGAACTGACTCGAATGTTTCAAAATAT 300
DB 241 GTAGGCGGAGGTGGGATTAATGATCTATTGAACTGACTCGAATGTTTCAAAATAT 300
QY 301 GCCACCATGACAGTGAACCGCTAAACCTCCGATTTATGTCGCGCAATACGGCGCT 360
DB 301 GCCACCATGACAGTGAACCGCTAAACCTCCGATTTATGTCGCGCAATACGGCGCT 360
QY 361 AATAAGCGCGCGCTGTTATATGACCGCATCTGATTCAGCGTAATGTCGCTGCTGCT 420
DB 361 AATAAGCGCGCGCTGTTATATGACCGCATCTGATTCAGCGTAATGTCGCTGCTGCT 420
QY 421 GGTTTGGCAACGACCGCGCTAACGATTA 456
DB 421 GGTTTGGCAACGACCGCGCTAACGATTA 456
```

RESULT 10
US-09-543-407-21

Sequence 21, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gpe3 of *Leishmania* major.
US-09-543-407-21

Query Match 88.1%; Score 401.6; DB 23; Length 456;
Best Local Similarity 92.5%; Pred. No. 8.8e-118;

Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCGCAATGCTGTTCTGCGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGCGCAATGCTGTTCTGCGCTGCTGCT 60
QY 61 GGGGTCTTCCACATATGGGCGGCGGCTGATCATACGGGCGGCAATGTTCCGGC 120
DB 61 GGGGTCTTCCACATATGGGCGGCGGCTGATCATACGGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 AGCGATGCCGCTGAAATCTGAAAAGCATTAACCGAGGCGGTTATGTTAAAGGCGCGAT 240
DB 181 AGCGATGCCGCTGAAATCTGAAAAGCATTAACCGAGGCGGTTATGTTAAAGGCGCGAT 240
QY 241 GTAGGCGGAGGTGGGATTAATGATCTATTGAACTGACTCGAATGTTTCAAAATAT 300
DB 241 GTAGGCGGAGGTGGGATTAATGATCTATTGAACTGACTCGAATGTTTCAAAATAT 300
QY 301 GCCACCATGACAGTGAACCGCTAAACCTCCGATTTATGTCGCGCAATACGGCGCT 360
DB 301 GCCACCATGACAGTGAACCGCTAAACCTCCGATTTATGTCGCGCAATACGGCGCT 360
QY 361 AATAAGCGCGCGCTGTTATATGACCGCATCTGATTCAGCGTAATGTCGCTGCTGCT 420
DB 361 AATAAGCGCGCGCTGTTATATGACCGCATCTGATTCAGCGTAATGTCGCTGCTGCT 420
QY 421 GGTTTGGCAACGACCGCGCTAACGATTA 456
DB 421 GGTTTGGCAACGACCGCGCTAACGATTA 456
```

RESULT 11
US-09-543-407-27

Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gpe3 of *Leishmania* major.
US-09-543-407-27

Query Match 88.1%; Score 401.6; DB 23; Length 456;
Best Local Similarity 92.5%; Pred. No. 8.8e-118;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCGCAATGCTGTTCTGCGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGCGCAATGCTGTTCTGCGCTGCTGCT 60
QY 61 GGGGTCTTCCACATATGGGCGGCGGCTGATCATACGGGCGGCAATGTTCCGGC 120
DB 61 GGGGTCTTCCACATATGGGCGGCGGCTGATCATACGGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 CCGGACTCAACGTTGAGCATTTATCGATACGCTGCTGCTGCTGCTGCTGCTGCT 180
```

Db 121 CCGAGCTCAGGTTGAGATTATATGATACGGTTCCGTAACGCTGCCCTTCTCTGCA 180
Qy 181 AGCGATGCCGTTAAATCTGAAAACGACATTCACAGAGCGGTTATGTATACGGCCGAT 240
Db 181 AGCGATGCCGTTAAATCTGAAAACGACATTCACAGAGCGGTTATGTATACGGCCGAT 240
Qy 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAACTGACAGAAAGTTTCAAGAAATAT 300
Db 241 TATGATCAGCTGGTTACCGCTGTGTGTTACCGTAAGGACATGCAATTCAGAAATAT 300
Qy 301 GCCACATCGACAGATGGAACGCTTAAAACTCCGATATTAATGTCGCGCAATACGGCCGT 360
Db 301 GCCACATCGACAGATGGAACGCTTAAAACTCCGATATTAATGTCGCGCAATACGGCCGT 360
Qy 361 AATAACGCGCGCTGGTTAATACAGCCCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Db 361 AATAACGCGCGCTGGTTAATACAGCCCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Qy 421 GGTTTGGCAACAACGCCAGCGCTAACAGTAATTA 456
Db 421 GGTTTGGCAACAACGCCAGCGCTAACAGTAATTA 456

RESULT 12
US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: Parseq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-29

Query Match 87.4%; Score 398.4; DB 23; Length 456;
Best Local Similarity 92.1%; Pred. No. 9.4e-117;
Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACAGATCGAGTTTCTGGAGTCTCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACAGATCGAGTTTCTGGAGTCTCTGGCT 60
Qy 61 GGGCTGTTCCACAATGGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
Db 61 GGGCTGTTCCACAATGGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGCTTGAAGATTTATCAAGTTCGCTTCCGTAACGTCGCTCTCTGCA 180
Db 121 CCGGACTCAAGCTTGAAGATTTATCAAGTTCGCTTCCGTAACGTCGCTCTCTGCA 180
Qy 181 AGCGATGCCGTTAAATCTGAAAACGACATTCACAGAGCGGTTATGTATACGGCCGAT 240
Db 181 AGCGATGCCGTTAAATCTGAAAACGACATTCACAGAGCGGTTATGTATACGGCCGAT 240
Qy 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAACTGACAGAAAGTTTCAAGAAATAT 300
Db 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAACTGACAGAAAGTTTCAAGAAATAT 300

Qy 301 GCCACATCGACAGATGGAACGCTTAAAACTCCGATATTAATGTCGCGCAATACGGCCGT 360
Db 301 GCCACATCGATGATGAGCTGGTTACCGCTGTGTGTTACCGTAAGGACATGCAAGCGCGT 360
Qy 361 AATAACGCGCGCTGGTTAATACAGCCCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Db 361 AATAACGCGCGCTGGTTAATACAGCCCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Qy 421 GGTTTGGCAACAACGCCAGCGCTAACAGTAATTA 456
Db 421 GGTTTGGCAACAACGCCAGCGCTAACAGTAATTA 456

RESULT 13
US-08-233-642A-54
Sequence 54, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1...357
US-08-233-642A-54

Query Match 72.8%; Score 331.8; DB 6; Length 361;
Best Local Similarity 99.4%; Pred. No. 2.2e-95;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 64 GTGCTTCACAATGGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGCCG 123
Db 1 GTGCTTCACAATGGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGCCG 60
Qy 124 GACTCAACGTTGAGCATTTTATCAAGTTCGCTTCCGTAACGTCGCTCTGCAAGC 183
Db 61 GACTCAACGTTGAGCATTTTATCAAGTTCGCTTCCGTAACGTCGCTCTGCAAGC 120

Search completed: March 18, 2004, 02:46:32Z
Job time : 2926.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 456
Sequence: 1 atggaacttttaaaagtgc.....ccacgctacacgtactataa 456

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents NA New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	254.4	55.8	675	US-09-741-873C-3	Sequence 3, Appli
2	210.4	46.1	396	US-09-741-873C-1	Sequence 1, Appli
3	33.4	7.3	3990	US-09-830-230A-629	Sequence 629, App
4	32.4	7.1	150223	US-10-624-149A-1	Sequence 1, Appli
5	31.4	6.9	1588	US-10-767-701-14522	Sequence 14522, A
6	30.8	6.8	1575	US-10-788-792-27	Sequence 27, Appli
7	29.6	6.5	24900	US-10-767-471-10736	Sequence 10736, A
8	29.4	6.4	300	US-10-767-701-28437	Sequence 28437, A
9	29.4	6.4	1355	US-10-045-674A-593	Sequence 593, App
10	29.4	6.4	1400	US-60-545-213-2134	Sequence 2134, App
11	29.4	6.4	1400	US-60-545-213-2135	Sequence 2135, App
12	29.4	6.4	1400	US-60-545-213-6406	Sequence 6406, App
13	29.4	6.4	1400	US-60-545-213-6407	Sequence 6407, App
14	29.2	6.4	249	US-10-771-241-117	Sequence 117, Appli
15	29.2	6.4	757	US-10-771-241-29	Sequence 29, Appli
16	28.8	6.3	83009	US-10-417-375A-143	Sequence 143, App
17	28.6	6.3	1186	PCT-US04-05654-1030	Sequence 1030, App
18	28.4	6.2	394468	US-60-548-091-5725	Sequence 5725, App
19	28.2	6.2	201	US-60-548-091-22688	Sequence 22688, A
20	28.2	6.2	596	US-10-767-701-5378	Sequence 5378, App
21	28.2	6.2	610	US-10-767-701-8781	Sequence 8781, App
22	28.2	6.2	1295	US-10-100-683-5480	Sequence 5480, App
23	28.2	6.2	1746	PCT-US04-05654-2555	Sequence 2555, App
24	28.2	6.2	3416	US-10-100-683-3397	Sequence 3397, App
25	28.2	6.2	8206	US-10-021-698A-3529	Sequence 3529, App
26	28.2	6.1	237	US-10-767-701-31417	Sequence 31417, A

c	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
c	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
c	30	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, App
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
c	32	27.6	6.1	600	7	US-60-545-213-6200	Sequence 2020, App
c	33	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, App
c	34	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
c	35	27.6	6.1	3035	6	US-10-786-892-279	Sequence 279, App
c	36	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, App
c	37	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, App
c	38	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, App
c	39	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	40	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, App
c	41	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
c	42	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, App
c	43	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, App
c	44	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli
c	45	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, App

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-064
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 06/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 06/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 55.8%; Score 254.4; DB 5; Length 675;
Best Local Similarity 72.4%; Pred. No. 2.5e-77;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 1 ATGAACTTTAAAGTGCAGCAATTCGATTTCTGCGAGTCTGAGCT 60
DB 83 ATGAACTTTAAAGTGCAGCAATTCGATTTCTGCGAGTCTGAGCT 142
QY 61 GCGTCGTTCCAAATGCGCGCGGCGGTAATATACGCGCGGCAATGTTCCGGC 120
DB 143 GGTGTTTCTCTCGTACGCGCGGCGGTAATATAGCGGC 202
QY 121 CCGACTCAACGTTGACATTTATACGAGGTTCCGTAACGCTGCGCTTCTGCA 180

```
Db 203 CCAATTCGAGCTGACATTTACGATGCGGTAACCTGCACTTGTCTGCAA 262
Qy 181 AGCGATGCCCTTAATCTGAAACGACATTAACCGAGCGGTTATGTTAAAGCGCGCAT 240
Db 263 ACTGATGCGCCCTTAACCTGACTGACTATTAATCCAGCATGCGCGGTTATGTTGAGAT 322
Qy 241 GTAGGCCGAGGTGGGATATAGTAACTGAACTGATGCTAGAAATGTTTCAAAATAT 300
Db 323 GTTGTGACGGGCTGAGATGACACTCAATGCAATGACCAAGTGTGCTTGGTAAACAGC 382
Qy 301 GCCACATGACCGCTGAGTGAACGCTTAAACTCGATATTAATGCTGCGCAATACGGCGGT 360
Db 383 GCTACTCTTGAATGAGTGAACGCGCAAAATTTGAAATGACGTTTAAACGTTTGGTGT 442
Qy 361 AATAACCGCGCTGCTTAAATGACAGCGCATGATTCAGCGTAAATGTTGCGTACGTT 420
Db 443 GGCACAGGTGCTGAGTTGACCAAGACTGACATCTTAACCTCTCCGTCACGTAAGTGT 502
Qy 421 GGTTTGGCAACAACGCGGCTTAAACAGTATTAA 456
Db 503 GGCCTTGTGTAACAACGCGGCTCACTCACTACTAA 538
```

RESULT 2

```
US-09-741-873C-1
Sequence 1, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
FILE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 396
TYPE: DNA
ORGANISM: Escherichia coli
US-09-741-873C-1
```

```
Query Match 46.1%; Score 210.4; DB 5; Length 396;
Best Local Similarity 70.7%; Pred. No. 2,2e-62;
Matches 280; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
```

```
Qy 61 GGCCTGCTTCCCAATGCGCGCGCGGCGGTAATCATTAACGCGCGCAATAGTCCGCGC 120
Db 1 GGTGTTCTTCTCAATGCGCGCGCGGTAACGCGGTGCGGTAAATATAGCGCGC 60
Qy 121 CCGGACTCAAGCTTGAGCATTTATGATCGGTTCCGTAAGCGTGGCTTCTTGGCAA 180
Db 61 CCAAAATTCGAGCTGAACATTTACGATGCGGTGCGGTAACTTGGCACTTGTCTTGGAA 120
Qy 181 AGGATGCCCCGTAATCTGAACGACATTAACCAAGCGGTTATGTTAAGCGCGCAT 240
Db 121 ACTGATGCGCGTAATCTGACTTATTAACCAAGCATGCGCGGTTATGTTGAGAT 180
Qy 241 GTAGGCCGAGGTGCGGATATAGTAACTGAACTGATGCTAGAAATGTTTCAAAATAT 300
Db 121 GGTGTTCTTCTCAATGCGCGCGCGGTAATCATTAACGCGCGCAATAGTCCGCGC 120
```

```
Db 181 GTTGTGAGGCTGATGACAGCTCAATGATCTGATCCCAAGGTGGCTTGGTAAACAGC 240
Qy 301 GCCACATGACCGAGTGAACGCTTAAACTCGATATTAATGCTGCGCAATACGGCGGT 360
Db 241 GCTACTCTTGAATGAGTGAACGCGCAAAATTTGAAATGACGTTTAAACGTTTGGTGT 300
Qy 361 AATAACCGCGCTGCTTAAATGACAGCGCATGATTCAGCGTAAATGTTGCGTACGTT 420
Db 301 GGCACAGGTGCTGAGTTGACCAAGACTGACATCTTAACCTCTCCGTCACGTAAGTGT 360
Qy 421 GGTTTGGCAACAACGCGGCTTAAACAGTATTAA 456
Db 361 GGCCTTGTGTAACAACGCGGCTCACTCACTACTAA 396
```

RESULT 3

```
US-09-830-230A-629
Sequence 629, Application US/09830230A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
FILE OF INVENTION: Lyme Disease Vaccines
FILE REFERENCE: PB481US
CURRENT APPLICATION NUMBER: US/09/830,230A
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: PCT/US98/12718
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/057,483
PRIOR FILING DATE: 1997-09-03
PRIOR APPLICATION NUMBER: 60/053,344
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/053,377
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/050,359
PRIOR FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 756
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 629
LENGTH: 3990
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1135)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1143)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1148)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1210)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1244)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1250)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1251)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1251)
OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629
```

Query Match 7.3%; Score 33.4; DB 5; Length 3990;
Best Local Similarity 47.4%; Pred. No. 0.24;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 157 GCTACGCTGGCTTCTCTGCAAGCGATGCCCTTAATCTGAACACCATTCACG 216
Db 3433 GTTACGGCTGGGGCTGCTGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAG 3492
Qy 217 AGCGGTATGTTACGCGCGCGCATTTAGCGCGGATTAATGACTATTGACTG 276
Db 3493 ATTGCTGCTGATTTAGGAGAGGATATGAGATGCTGCGGATTTTGTAGAGATGAGATG 3552
Qy 277 ACTGCAATGGTTTCAGAAATATGCAATCCATCCAGCAGTGGAGCGTAAATCCGAT 336
Db 3553 AAGAGGATGATCAGATTGCTGCTCTCTATTGCTTTGAGGGGAGATGCTTAAGAGATGAAAG 3612
Qy 337 ATTACTGTGGCCCATATACGGCGGTAAATACG 367
Db 3613 TTTGCTGTGAAGAGATGATGATGTGAGAAAG 3643

RESULT 4
US-10-624-149A-1
Sequence 1, Application US/10624149A
GENERAL INFORMATION:
APPLICANT: Neubauer, Antonie
APPLICANT: Ziegler, Christina
TITLE OF INVENTION: gm-Negative EHV-Mutants without Heterologous Elements
FILE REFERENCE: 1/1372
CURRENT FILING DATE: 2003-07-21
CURRENT APPLICATION NUMBER: US/10/624,149A
PRIOR FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: 60/403,282
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: DE 10233064
PRIOR FILING DATE: 2003-04-11
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Version 3.0
SEQ ID NO 1
LENGTH: 150223
TYPE: DNA
ORGANISM: Equine herpesvirus 1
PUBLICATION INFORMATION:
AUTHORS: Teiford, E.A.
AUTHORS: Watson, M.S.
AUTHORS: McBride, K.
AUTHORS: Davison, A.J.
TITLE: The DNA sequence of equine herpesvirus-1
JOURNAL: Virology
VOLUME: 189
ISSUE: 1
PAGES: 304-316
DATE: JUL-1992
DATABASE ACCESSION NUMBER: NC 001491, NCBI
DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 2.8;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 173 CTCTGCAAGAGATGCCCTTAATCTGAAGACCATTAACCAAGCGGTTATGTAACG 232
Db 95342 CTGTGTGTAACGCTGGGCTTGTGTTGATGACGGCTCATTTCCACAGCATGTTTCCTACG 95601
Qy 233 GCGCGCATGTAGCGCAGGCTGCGGATATAGTACTATGAGTCACTAGTCAAGATGTTTCA 292
Db 95602 TCAAGCCCATATAGCCACGACGAGATGCTTCTGTATACCAATGCTCTGGGGCGCGTCA 95661
Qy 293 GAATTAATGCCACCATCGACCACTG 318

Db 95662 GAGAGCTTATTCACACGAGCTGTTG 95687

RESULT 5
US-10-767-701-14522
Sequence 14522, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(5335)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 14522
LENGTH: 1588
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

Query Match 6.9%; Score 31.4; DB 6; Length 1588;
Best Local Similarity 44.5%; Pred. No. 0.76; Indels 156; Gaps 0;
Matches 125; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy 40 GTTCTGGCAGTGTCTGCTGGCGGCTTCCACAATGGGCGCGGCTAATCATAC 99
Db 455 GGTGTGGCGGCTTCCGAGTGTGTGCTATGAGCGGCGGCGGCTATGAGTGGC 514
Qy 100 GCGCGCGCATATTTCCGCGCGGACTCAACGTTAGCATTTATCACTACGTTCCGCT 159
Db 515 GCGCGTGTATGT 574
Qy 160 AACCTGCGCTTGTCTGCAAGGATGCCGTAATCTGAAGACCATTAACCGACG 219
Db 575 GGT 634
Qy 220 GGTATGTATACGCGCGCATGTAGGCGCAGGCTGCGATTAATGATTAATTAATTAATTAAT 279
Db 635 GTATGTAACAGCTTGTGCAAGAGTACTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 694
Qy 280 CAGATGTTTCAAAATTAATCCACCATGACCATGAGTGA 320
Db 695 GCTGT 725

RESULT 6
US-10-788-792-27/c
Sequence 27, Application US/10788792
GENERAL INFORMATION:
APPLICANT: Bayer Pharmaceuticals Corporation
APPLICANT: Eveleigh, Deepa
TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
FILE REFERENCE: 5152
CURRENT APPLICATION NUMBER: US/10/788,792
CURRENT FILING DATE: 2004-02-27
PRIOR APPLICATION NUMBER: US 60/450,655
PRIOR FILING DATE: 2003-02-28
NUMBER OF SEQ ID NOS: 254
SOFTWARE: PatentIn version 3.2
SEQ ID NO 27
LENGTH: 1575
TYPE: DNA
ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.2;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```

/ GENERAL INFORMATION: Application US/10767471
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CLO01505
/ CURRENT APPLICATION NUMBER: US/10/767,471
/ CURRENT FILING DATE: 2004-01-30
/ NUMBER OF SEQ ID NOS: 50231
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10736
/ LENGTH: 24900
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(24900)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Table
US-10-767-471-10736

```

Qy	207	CATTACCAAGAGCGTTATGTATACGGCGCATGTAGCCAGGGTCCGATATATGTAC	266
Db	10048	CAGATCTCAGCTGCTTTTATGTGTCTAAAGTCCGTGGTCAATGTGTGATATATGTG	9983
Qy	267	TATTGACGTACTCAGATGAGTTTATGAAATATATGACCATGTACCAAGTGAAGCTTA	326
Db	9988	TCTCTGAACCCAAACTGTATGTACTATAAATAATATGCAATCTGTGTTGTATATTATA	9929
Qy	327	AAAC 330	
Db	9928	AATC 9925	

? Sequence 28437, Application US/10767701
 ? GENERAL INFORMATION:
 ? APPLICANT: Kowallik, David K.
 ? APPLICANT: Zhou, Yihua
 ? APPLICANT: Cao, Yongwei
 ? TITLE OF INVENTION: Nucleic Acid Molecule
 ? TITLE OF INVENTION: Plants and Uses Thereof
 ? FILE REFERENCE: 38-21(53535)B
 ? CURRENT APPLICATION NUMBER: US/10/767,701
 ? CURRENT FILING DATE: 2004-01-29
 ? NUMBER OF SEQ ID NOS: 63128
 ? SEQ ID NO 28437
 ? LENGTH: 300
 ? TYPE: DNA
 ? ORGANISM: Sorghum bicolor
 ? FEATURE:
 ? OTHER INFORMATION: Clone ID: 7551688
 US-10-767-701-28437

```
Query Match      6.4%; Score 29.4; DB 6; Length 300;
Best Local Similarity 54.1%; Pred. No. 1.7;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0
```

[illegible]

```

sequence 993, Application US/10045674A
GENERAL INFORMATION:
APPLICANT: LADNER, ROBERT C.
APPLICANT: COHEN, EDWARD H.
APPLICANT: NASTRI, HORACIO G.
APPLICANT: ROCKEY, KRISTIN L.
APPLICANT: HOET, RENE
APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
FILE REFERENCE: DMAX/002 C1P2
CURRENT APPLICATION NUMBER: US/10/045,674A
CURRENT FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: 06/198,069
PRIOR FILING DATE: 2000-04-17
PRIOR APPLICATION NUMBER: 09/837,306
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 635
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 593
LENGTH: 1355
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: M13-III
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1305)
US-10-045-674A-593

```

Query Match	6.4%;	Score 29.4;	DB 6;	Length 1355;
Best Local Similarity	25.2%;	Pred. No. 3.4;		
Matches 39;	Conservative 50;	Mismatches 66;		

QY	78	GAGCGGCGCGCGGTAATCAATACCGCGCGCGGCGCATAGTTCGGCCGCACTCAACGTTGAG	137
Db	768	KGCGYGGYGYMSYGGYGGMGYMSYGGYGGYMSYGARGGYGGYGGYMSYGARCGMGCG	827
QY	138	CATTATACGTAACGTTCCGCTAACCGCTCGCTTGCCTGCACAAAGCATGCCCTTAATC	197
Db	828	YGYMSYGGYGGYGGYMSYGGYMSYGGYGYATTYGAATYAGRAARAATGGCAATYGCYAA	887
QY	198	TAAACGACCATTAACCGACGACGGTATGGTAACG	232
Db	888	YAAAGSGCATGACTACGARAATCTCAATARAATG	922

```

? Sequence 2134, Application US/60545213
? GENERAL INFORMATION:
?
? APPLICANT: Wyeth
?
? APPLICANT: Mounts, William Martin
?
? TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
?
? FILE REFERENCE: 1a201083 (031896-042099)
?
? CURRENT APPLICATION NUMBER: US/60/545,213
?
? CURRENT FILING DATE: 2004-02-18

```


QY 212 CCCAGAGCCGGTATATGGTAAACGGCGCCCATGTATAGGCGAAGGGGACGGAATTAATAGTACTATTG 271

Db 437 CCTGGGCGCATAGCGTTTCGATGACAGGTTGACCCGACGTTGGGGTGGATTCGGGTTATTGA 378

QY 272 AACTGACTCAGAAATGCTTTCAGAAATAATATGTCACCATTCGACAGATGGAAACGCTTAAATACT 331

Db 377 TTGGGGTTCTCCATCAGCTGCTTGGAAAGGCGACAAGACCATCTTCCTCAGACTCTGGAACT 318

QY 332 CGGATTTACTCTCGGCGCAATACGGCGGGTATTAACGCGCGCTGGTGAATCAGACCGCAT 391

Db 317 CCGTTGAGACAGGCCGCTGCTGATGGCATTTAACTCGCTGTACTGGTCCATTCAGGTTTTCT 258

QY 392 CTGATTTCCAGCGTAATGGTGCCTCAGGTTGGTTTTGGCAACACGCAACGCGCTTAACAG 450

Db 257 GGGATCTCCCTCCAGTTGTTGATATTTCGAGTGAATGAAGGGGTGTGACTTG 199

RESULT 14

US-10-771-241-117/c
 ; Sequence 117, Application US/10771241
 ; GENERAL INFORMATION:
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Forsyth, R. Allen
 ; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
 ; FILE REFERENCE: ELITRA.001C1
 ; CURRENT APPLICATION NUMBER: US/10/771,241
 ; PRIOR FILING DATE: 2004-02-03
 ; PRIOR APPLICATION NUMBER: 09/492,709
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: 60/117,405
 ; NUMBER OF SEQ ID NOS: 485
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 117
 ; LENGTH: 249
 ; TYPE: DNA
 ; ORGANISM: E. Coli
 ; US-10-771-241-117

Query Match 6.4%; Score 29.2; DB 6; Length 249;
 Best Local Similarity 52.5%; Pred. No. 1.8;

Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY 295 AATAATGCCACCATGACCAAGTGAACCGTAAACCTCCGATATTACTGTGCGCAATAC 354
 Db 180 AAAATTGAAGCCATCGACTTTACCAAGCCAAACGCGTGTGATCCATCCGCCGACTAC 121
 OY 355 GCGGTATATACGCGCGCTGTATATACGCCGATCTGATTCAGCGTATAGTGCCT 414
 Db 120 GGCACCGACTATCCCGAGAGATAGTCAATGAATCCACCTCATCTTACCTGGCAT 61
 OY 415 CA 416
 Db 60 GA 59

RESULT 15

US-10-771-241-29
 ; Sequence 29, Application US/10771241
 ; GENERAL INFORMATION:
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Forsyth, R. Allen
 ; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
 ; FILE REFERENCE: ELITRA.001C1
 ; CURRENT APPLICATION NUMBER: US/10/771,241
 ; PRIOR FILING DATE: 2004-02-03
 ; PRIOR APPLICATION NUMBER: 09/492,709
 ; PRIOR FILING DATE: 2000-01-27
 ; PRIOR APPLICATION NUMBER: 60/117,405
 ; NUMBER OF SEQ ID NOS: 485
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 29
 ; LENGTH: 757
 ; TYPE: DNA
 ; ORGANISM: E. Coli
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)...(757)
 ; OTHER INFORMATION: n = A,T,C or G
 ; US-10-771-241-29

Query Match 6.4%; Score 29.2; DB 6; Length 757;
 Best Local Similarity 52.5%; Pred. No. 3.1;
 Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

OY 295 AATAATGCCACCATGACCAAGTGAACCGTAAACCTCCGATATTACTGTGCGCAATAC 354
 Db 264 AAAATTGAAGCCATCGACTTTACCAAGCCAAACGCGTGTGATCCATCCGCCGACTAC 323
 OY 355 GCGGTATATACGCGCGCTGTATATACGCCGATCTGATTCAGCGTATAGTGCCT 414
 Db 324 GGCACCGACTATCCCGAGAGATAGTCAATGAATCCACCTCATCTTACCTGGCAT 383
 OY 415 CA 416
 Db 384 GA 385

Search completed: March 17, 2004, 08:25:30
 Job time : 71.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds

(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-2

Perfect score: 456
Sequence: 1 atgaataaacattgtctatc.....ctcgctcaccacgaagctaa 456Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA Main:

```
1: /cgn2_6/ptodata/2/pna/PCRTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US092b_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US093b_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US093b_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095b_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095b_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US095c_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US095d_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096a_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096a_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096c_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US096b_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US096b_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097b_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US097b_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US097c_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US098a_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US098b_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US098c_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US098b_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US099a_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US099b_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US099c_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US099d_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US099b_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US099f_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US100a_COMB.seq.*
```

```
44: /cgn2_6/ptodata/2/pna/US100b_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US101a_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US101b_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US102a_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US102b_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US103a_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US103b_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US104a_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US104b_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US107a_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US107b_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US6023a_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US6023b_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

US-09-543-407-2									
Query Match	100.0%	Score 456	DB 23	Length 456					
Best Local Similarity	100.0%	Pred. No. 4.9e-130							
Matches 456	Conservative 0	Mismatches 0	Indels 0	Gaps 0					
QY	1	ATGAAAAACAATTTGTTATTATTGATGATGTGACAAATACCTGGGTGGCCCTGGGATTTGCCAAC	60						
Db	1	ATGAAAAACAATTTGTTATTATTATGATGTGACAAATACCTGGGTGGCCCTGGGATTTGCCAAC	60						
QY	61	GGGCAAAATTAATGATCTGGCTGCTTCAGAAATATAATTTGGGTAAATGAATTAAAGCAAG	120						
Db	61	GGGCAAAATTAATGATCTGGCTGCTTCAGAAATATAATTTGGGTAAATGAATTAAAGCAAG	120						
QY	121	TCTTCATTTAATACGAGGGGCCATTATTGTCMAATCGGACGGAATATAGTCCAGAGTA	180						
Db	121	TCTTCATTTAATACGAGGGGCCATTATTGTCMAATCGGACGGAATATATAGTCCAGAGTA	180						
QY	181	CGCCAGAGAGATATAAACTATTGTCCGTTATTTTCACAAAGAGAGGAATATATCGGCG	240						
Db	181	CGCCAGAGAGATATAAACTATTGTCCGTTATTTTCACAAAGAGAGGAATATATCGGCG	240						
QY	241	AAATGTGACCGAGGAGGAAATTAATTAACCTTGGCGTATATTGACGAAACGGGCAATGCCAAC	300						
Db	241	AAATGTGACCGAGGAGGAAATTAATTAACCTTGGCGTATATTGACGAAACGGGCAATGCCAAC	300						
QY	301	GATCCAGATATATGCAAAAGCGCTTACGTTAATATGTCAGCTATTATTCAGAAAGTTCT	360						
Db	301	GATCCAGATATATGCAAAAGCGCTTACGTTAATATGTCAGCTATTATTCAGAAAGTTCT	360						
QY	361	GGAAATTAAGGCAATATTAACCCAGTAGCGTACGAGAGAAACAGAGTTGTATGTCAGAAA	420						
Db	361	GGAAATTAAGGCAATATTAACCCAGTAGCGTACGAGAGAAACAGAGTTGTATGTCAGAAA	420						
QY	421	CAGTCGATATGGCTATTGCGTACGCCAACGCTAA	456						
Db	421	CAGTCGATATGGCTATTGCGTACGCCAACGCTAA	456						
RESULT 2									
US-09-543-407-4									
; Sequence 4, Application US/09543407									
; GENERAL INFORMATION:									
; APPLICANT: White, Aaron P.									
; APPLICANT: Doran, James L.									
; APPLICANT: Collinson, S. Karen									
; APPLICANT: Kay, William W.									
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES									

```
Db 121 TCTTATTATACAGGAGCCATATGCTGACGCTGGGACATATATATATGCTCAGTTA 180
Qy 181 CGCCAGGAAGATCAAACTATGTCCTGTTATTTCAACAAGAGAAATATATGCGGCG 240
Db 181 CGCCAGGAAGATCAAACTATGTCCTGTTATTTCAACAAGAGAAATATATGCGGCG 240
Qy 241 AAAGTCGACGAGGAGGAAATATATGCTGTTATTTGAGCAAAACGGGCAATGCCAAC 300
Db 241 AAAGTCGACGAGGAGGAAATATATGCTGTTATTTGAGCAAAACGGGCAATGCCAAC 300
Qy 301 GATGCGCATATATCGCAAAAGCGCTTACGCTATAGTGAAGCTATTATCCAGAAAGTTCT 360
Db 301 GATGCGCATATATCGCAAAAGCGCTTACGCTATAGTGAAGCTATTATCCAGAAAGTTCT 360
Qy 361 GGAATTAAGCCCAATATTAACCAAGTACGAGTACGCAAAACAGCAGTTGTAAGTCAAAA 420
Db 361 GGAATTAAGCCCAATATTAACCAAGTACGAGTACGCAAAACAGCAGTTGTAAGTCAAAA 420
Qy 421 CAGTCGATATGCTATTCGCTGCTCAACCAAGCTAA 456
Db 421 CAGTCGATATGCTATTCGCTGCTGACACACGTTAA 456
```

```
RESULT 3
US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; PRIORITY FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-171
```

```
Query Match 64.7%; Score 295; DB 17; Length 561;
Best Local Similarity 78.0%; Pred. No. 3.9e-80;
Matches 355; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 1 ATGAAAAACAATGTTATTTATGATGTTGACAAATACGCTGGCGCTGGGATTGCAAC 60
Db 106 ATGAAAAACAATGTTATTTATGATGTTGACAAATACGCTGGCGCTGGGATTGTAATC 165
Qy 61 GCGACAATTAATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Db 166 GCGACAATTAATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
Qy 121 TCTTATTATATCAGGAGGCGCATTTATGTCAGTCAAGTCGCAACGATATATAGTCCAGATA 180
Db 226 GCTTATATCATATCAGGAGGCGCATTTATGTCAGTCAAGTCAAGTCAAGTCAAGTCAAGT 285
Qy 181 CGCCAGGAAGATCAAACTATGTCCTGTTATTTCAACAAGAGAAATATATGCGGCG 240
Db 286 CGCCAGGAAGATCAAACTATGTCCTGTTATTTCAACAAGAGAGGCGGGAATATATGCGGCG 345
Qy 241 AAAGTCGACGAGGAGGAAATATATGCTGTTATTTGAGCAAAACGGGCAATGCCAAC 300
Db 346 AAAGTCGACGAGGAGGAAATATATGCTGTTATTTGAGCAAAACGGGCAATGCCAAC 405
Qy 301 GATGCGCATATATCGCAAAAGCGCTTACGCTATAGTGAAGCTATTATCCAGAAAGTTCT 360
Db 406 GATGCGCATATATCGCAAAAGCGCTTACGCTATAGTGAAGCTATTATCCAGAAAGTTCT 465
Qy 361 GGAATTAAGCCCAATATTAACCAAGTACGAGTACGCAAAACAGCAGTTGTAAGTCAAAA 420
Db 466 GGAATTAAGCCCAATATTAACCAAGTACGAGTACGCAAAACAGCAGTTGTAAGTCAAAA 525
```

```
Qy 421 CAGTCGATATGCTATTCGCTGCTCAACCAAGCTAA 455
Db 526 CAGTCGATATGCTATTCGCTGCTGACACGTTAA 560
```

```
RESULT 4
US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIORITY FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-171
```

```
Query Match 64.7%; Score 295; DB 17; Length 561;
Best Local Similarity 78.0%; Pred. No. 3.9e-80;
Matches 355; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 1 ATGAAAAACAATGTTATTTATGATGTTGACAAATACGCTGGCGCTGGGATTGCAAC 60
Db 106 ATGAAAAACAATGTTATTTATGATGTTGACAAATACGCTGGCGCTGGGATTGTAATC 165
Qy 61 GCGACAATTAATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
Db 166 GCGACAATTAATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 225
Qy 121 TCTTATTATATCAGGAGGCGCATTTATGTCAGTCAAGTCGCAACGATATATAGTCCAGATA 180
Db 226 GCTTATATCATATCAGGAGGCGCATTTATGTCAGTCAAGTCAAGTCAAGTCAAGTCAAGT 285
Qy 181 CGCCAGGAAGATCAAACTATGTCCTGTTATTTCAACAAGAGAAATATATGCGGCG 240
Db 286 CGCCAGGAAGATCAAACTATGTCCTGTTATTTCAACAAGAGAGGCGGGAATATATGCGGCG 345
Qy 241 AAAGTCGACGAGGAGGAAATATATGCTGTTATTTGAGCAAAACGGGCAATGCCAAC 300
Db 346 AAAGTCGACGAGGAGGAAATATATGCTGTTATTTGAGCAAAACGGGCAATGCCAAC 405
Qy 301 GATGCGCATATATCGCAAAAGCGCTTACGCTATAGTGAAGCTATTATCCAGAAAGTTCT 360
Db 406 GATGCGCATATATCGCAAAAGCGCTTACGCTATAGTGAAGCTATTATCCAGAAAGTTCT 465
Qy 361 GGAATTAAGCCCAATATTAACCAAGTACGAGTACGCAAAACAGCAGTTGTAAGTCAAAA 420
Db 466 GGAATTAAGCCCAATATTAACCAAGTACGAGTACGCAAAACAGCAGTTGTAAGTCAAAA 525
Qy 421 CAGTCGATATGCTATTCGCTGCTCAACCAAGCTAA 455
Db 526 CAGTCGATATGCTATTCGCTGCTGACACGTTAA 560
```

```
RESULT 5
US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10/417,886
; PRIORITY FILING DATE: 2003-04-17
```

; PRIOR APPLICATION NUMBER: US/09/252,691C
 ; PRIOR FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,145
 ; PRIOR FILING DATE: 1998-07-24
 ; PRIOR APPLICATION NUMBER: US 60/074,787
 ; PRIOR FILING DATE: 1998-02-18
 ; NUMBER OF SEQ ID NOS: 11326
 ; SEQ ID NO 171
 ; LENGTH: 561
 ; TYPE: DNA
 ; ORGANISM: Enterobacter cloacae
 ; US-10-417-886-171

Query Match 64.7% Score 295; DB 51; Length 561;
 Best Local Similarity 78.0%; Pred. No. 3,9e-80;
 Matches 355; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTGTTATTTATGATGTTGACATATCTGGGTGGCGCTGGGATTCACAC 60
 DB 106 ATGAAAAACAACGTTGTTATGATGTTTACATTAATGAGTGGCGCTGGGTTGTAATC 165
 QY 61 GCGACAAATTATGATCTGGCTGCTGCAATATATATTTGGGTAATGATTAAGCAAG 120
 DB 166 GCAGTGATTCAGATTATGACAGTCTGAAATATATTTTCCGATTAATGATTAAGTAA 225
 QY 121 TCTTCAATTATATCGCGCGCATTTATGTCAGTGGCAGCGGATTAATGTCGCAAGTA 180
 DB 226 GCTTCAATCAATCAGGAGCATTTATGTCACAGGCTTCAAGAAATATTTCTGATGTA 285
 QY 181 GCGCAGGAGGATCAAAACTATTTGCTGTTATTTTCAAGAAAGAAATATGCGGCG 240
 DB 286 GCGCAGGAGGCTTCAAAATTTGCTGCTGTTATTTTCAAGAGGCGGGAATATCCGCGC 345
 QY 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGCTATTTGAGCAACCGGCAATTCGATTA 300
 DB 346 AATGTTGATCAGTCAGGAGCATTAATCTTGTATATGATCAGACCGGCAAGCGCAAC 405
 QY 301 GATGCAAGATATCGGAAAGCGCTTACGTTAATGATGAGCTATTTATCCAGAAAGTTCT 360
 DB 406 GATGCAAGATTAAGGAGGCGCTTATGCAACCGGCAATGATTAATCAAAAGGCTCG 465
 QY 361 GGAATATAGGCAATATTAACCGATACGTTACGCAAGAAAGCAAGATTTGTCAGAAA 420
 DB 466 GGTAAACAGGCGCAATATTAACGATATGTCAGCAAGAAAGCAAGATTTGTCAGAGA 525
 QY 421 CAGTGCATATGCTATTCGCGTCAACCAACGCTA 455
 DB 526 CAGTGCATATGCTATTCGCGTATTCACGCTTA 560

RESULT 6
 ; US-10-146-492B-80
 ; Sequence 80, Application US/10146492B
 ; GENERAL INFORMATION:
 ; APPLICANT: MMG Biotech AG
 ; TITLE OF INVENTION: Biochip
 ; FILE REFERENCE: REN-14043
 ; CURRENT APPLICATION NUMBER: US/10/146,492B
 ; CURRENT FILING DATE: 2002-05-15
 ; PRIOR APPLICATION NUMBER: EP 0111279.5
 ; PRIOR FILING DATE: 2001-05-17
 ; NUMBER OF SEQ ID NOS: 12816
 ; SEQ ID NO 80
 ; LENGTH: 100
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli K-12 MG1655
 ; FEATURE:
 ; OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
 ; PUBLICATION INFORMATION:
 ; RELEVANT RESIDUES: 39-58
 ; US-10-146-492B-80

Query Match 16.3%; Score 74.4; DB 45; Length 100;

Best Local Similarity 84.0%; Pred. No. 4.7e-12;
 Matches 84; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 QY 266 ACTTGCCATATTTAGCAAAAGGCAATGCCAAGTCAGTATATCGCAAGCGCTT 325
 DB 1 ACTTGCAATATATTTAGCAAGGCGGAGTCGCAACGATATTTGCAAGGTCCTT 60
 QY 326 ACGTAATAGTGCAGTATTTATCCAGAAAGTCTGGAAA 365
 DB 61 ATGTAATATCTGCATGATTTATCCAGAAAGTCTGGTAA 100

RESULT 7
 ; US-10-146-492B-82
 ; Sequence 82, Application US/10146492B
 ; GENERAL INFORMATION:
 ; APPLICANT: MMG Biotech AG
 ; TITLE OF INVENTION: Biochip
 ; FILE REFERENCE: REN-14043
 ; CURRENT APPLICATION NUMBER: US/10/146,492B
 ; CURRENT FILING DATE: 2002-05-15
 ; PRIOR APPLICATION NUMBER: EP 0111279.5
 ; PRIOR FILING DATE: 2001-05-17
 ; NUMBER OF SEQ ID NOS: 12816
 ; SEQ ID NO 82
 ; LENGTH: 100
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli K-12 MG1655
 ; FEATURE:
 ; OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
 ; PUBLICATION INFORMATION:
 ; RELEVANT RESIDUES: 57-76
 ; US-10-146-492B-82

Query Match 14.6%; Score 66.4; DB 45; Length 100;
 Best Local Similarity 79.0%; Pred. No. 1.4e-09;
 Matches 79; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 123 TTCATTTAATGCGCGCATTTATGTCAGAGTCGACGATTAATATGTCGCAAGTACG 182
 DB 1 TTCATTTAATGCGCGCATTTATGTCAGAGTCGACGATTAATATGTCGCAAGTACG 60
 QY 183 CCAGAGAGATCAAACTATTTGCTGTTATTTCAAGAA 222
 DB 61 GCGAGAGGCTCAAACTTTTGGCGGTTGTCGCAAGAA 100

RESULT 8
 ; US-10-146-492B-81
 ; Sequence 81, Application US/10146492B
 ; GENERAL INFORMATION:
 ; APPLICANT: MMG Biotech AG
 ; TITLE OF INVENTION: Biochip
 ; FILE REFERENCE: REN-14043
 ; CURRENT APPLICATION NUMBER: US/10/146,492B
 ; CURRENT FILING DATE: 2002-05-15
 ; PRIOR APPLICATION NUMBER: EP 0111279.5
 ; PRIOR FILING DATE: 2001-05-17
 ; NUMBER OF SEQ ID NOS: 12816
 ; SEQ ID NO 81
 ; LENGTH: 100
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli K-12 MG1655
 ; FEATURE:
 ; OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
 ; PUBLICATION INFORMATION:
 ; RELEVANT RESIDUES: 38-57
 ; US-10-146-492B-81

Query Match 14.2%; Score 64.8; DB 45; Length 100;
 Best Local Similarity 76.0%; Pred. No. 4.4e-09;
 Matches 78; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

RESULT 12
US-09-739-449-208
; Sequence 208, Application US/09739449

Query Match	9.7%	Score 44.2	DB 31	Length 18792
Best Local Similarity	47.6%	Pred. No. 0.14		
Matches 130; Conservative	0	Mismatches 143	Indels 0	Gaps 0

RESULT 13
US-09-803-110-208

Query Match

QY 385 TACGGTACGCAGAAACAGCAGTTGTAGTCAG 417
||| | | | | | | | | |
Db 137348 GACGGCAATGCAACATTGCTGCAGCGTTGAG 137380

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```


US-08-466-194-14

Query Match 9.0%; Score 41; DB 8; Length 7218;

Best Local Similarity 3.0%; Pred. No. 0.44; Matches 11; Conservative 200; Mismatches 150; Indels 0; Gaps 0;

```
OY 63 GACAAATATGATCTGGCTTCGATATATATTTGCGTAATGATTAAGCAAGTC 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1454 GAGATGAAAGATTGTCACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395

OY 123 TTCATTATACGCGCCATTATGTCAGTCGCAAGATATAGTCAGAGTACG 182
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1394 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1335

OY 183 CCAGGAAGATCAAAACTATTGTCGTTATTTACAGAGAGGAATATCGGCGCA 242
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1334 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1275

OY 243 AGTCAGCAGGAGGAAATTAATCTTGCATATATGAGCAAGCGCAATGCCAACA 302
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1274 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1215

OY 303 TGCCAGTATATCGCAAGCGCTTACGTAATAGTCAGCTATTATCAGAAAGTTCTG 362
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1214 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1155

OY 363 AAATAGGCCAATATATCCAGTAGCGTACGAGCAAGAAAGAGGTTAGTCAGAAACA 422
    : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1154 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1095

OY 423 G 423
DB 1094 R 1094
```

RESULT 15

US-09-543-407-13

Sequence 13, Application US/09543407

GENERAL INFORMATION:

APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL, FIMBRIAL, SYSTEM FOR

TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543,407

CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13

LENGTH: 456

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA

OTHER INFORMATION: sequence containing the replacement fragment

OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-13

Query Match 8.2%; Score 37.6; DB 23; Length 456;

Best Local Similarity 51.2%; Pred. No. 2; Matches 88; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

```
OY 272 CGTATTATGACCAACGGGCAATGCCAGATGCCAGTATATCGCAAGCGCTTACGCTA 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 203 CGACCATTAACCCAGAGCGGTATGTAACGGCGCGATGAGCCAGGTCGGATATATA 262

OY 332 ATAGTCAGCTATTATCCAGAAAGTTCTGGAATAAGCCCAATATTAACCAATACGTA 391
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 263 GTACTATTGAATGACTGCTAGATGTTTCAGAAATTAATGCAACATCGACCATGGAACG 322

OY 392 CGCAGAAAAAGCAGTTGATGAGAAACAGTCGATATGCTATTGCGCT 443
```

DB 323 CTAAAACTCGATATTACTCTCGGCCAATATGATCAGCTGTTACCGCTGT 374

Search completed: March 18, 2004, 02:46:45
Job time : 2935.24 secs

THIS PAGE BLANK (USPTO)

RESULT 2

US-10-765-790-79/c
; Sequence 79, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; PRIOR FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: Patent version 3.2
; SEQ ID NO 79
; LENGTH: 401433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-79

Query Match
Best Local Similarity 7.4%; Score 33.8; DB 6; Length 401433;
Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 5 AAACAATGTTATTTATGATGTTGCAATGCTGGGCTGGGATTCGACCGGCA 64
DB 270713 AGAAATATTTTCTTTTCTTTTGTGATGTTTCTGAAACTCGGAAACAAACATGT 270654
QY 65 CAATATATGATCTGCTGCTGCTGCAATATATTTTGGGTTAATGATTAAGCAAGTCTT 124
DB 270653 CAATGAACTAATGCTTCCCTCGAAATATGTTTGAAGAAATATTAACCCCAAAAT 270594
QY 125 CATTATATGAGGGGCGCATTTATGTCAGTCGACGAGATTAATA 169
DB 270593 ACTAAGAAACGAAACAGATGATGATGTAATGGAACATAGAAATA 270549

RESULT 3

US-10-779-543-9321
; Sequence 9321, Application US/10779543
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; PRIOR FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114

PRIOR FILING DATE: 1998-03-31

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9321

LENGTH: 418
TYPE: DNA

ORGANISM: Homo sapiens
US-10-779-543-9321

Query Match

Best Local Similarity 7.2%; Score 32.8; DB 6; Length 418;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 87 AGAATATATTTTGGCGTAATGATTAAGCAAGTCTTATTTATTCAGCGGCGCATTTAT 146
DB 208 AGAATTCAGACTGGGGCCATGACCAAGGCAATTCGATCCTCAAGATTAATTGGA 267
QY 147 TGTCAAGTCGCGCAGGATTAATGTCCTCCAGAGTACGCCAGAGGATCAAACTATTGTC 206
DB 268 AGATTGATTTAGCAGTGTGCTTGAAACAGACTGGGCCAGGAGGCGTATTAAGCAGGAG 327
QY 207 CGTTATTTACA 218
DB 328 AGTCATTTAAA 339

RESULT 4

US-10-767-471-8836/c
; Sequence 8836, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: Cargill, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; PRIOR FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8836
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471-8836

Query Match

Best Local Similarity 7.1%; Score 32.4; DB 6; Length 201;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

QY 326 ACCGTATAGTGAAGTATTTCCAGAAAGTTCTGAAATTAAGGCCAATATTAACCCAGT 385
DB 142 ATGGAGATATGAGGGGGAATTCAGAGATGATGATTAATTAAGGAGCTATTTCCAGG 83
QY 386 ACCGTACGCAAAACAGCAGTGTAGTCAGAAACAGTCGCATATGG 433
DB 82 AGAATATCATGTATCATGTATGAGGACAGCAGAAACAGAAAGAGCTG 35

RESULT 5

US-10-767-471-34493/c
; Sequence 34493, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: Cargill, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; PRIOR FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34493
; LENGTH: 201
; TYPE: DNA

ORGANISM: Homo sapiens
US-10-767-471-34493

Query Match 7.1%; Score 32.4; DB 6; Length 201;
Best Local Similarity 55.6%; Pred. No. 0.78;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

Qy 326 ACGGTAATAGTGCAGCTATTATCCAGAAAGTTCTGAAATPAGGCCCAATTATCCCACT 385
Db 142 ATGGAGATGAGGCGGAGAAATCCAAATGATGATATATPAGAGGCTATTTCACAG 83

Qy 386 ACGGTACCCAGAAAAACAGCAGTTGTAGTCAGAAAAAGTCGCATATG 433
Db 82 AGGATTATCATGTACATGATGAGCAGAGCAGAAACAGAAAAAGCTGG 35

RESULT 6
US-10-767-471-501/c
Sequence 501, Application US/10767471
GENERAL INFORMATION:

APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CL001505
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 501
LENGTH: 1227
TYPE: DNA
ORGANISM: Homo sapiens
US-10-767-471-501

Query Match 7.1%; Score 32.4; DB 6; Length 1227;
Best Local Similarity 55.6%; Pred. No. 1.3;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

Qy 326 ACGGTAATAGTGCAGCTATTATCCAGAAAGTTCTGAAATPAGGCCCAATTATCCCACT 385
Db 476 ATGGAGATGAGGCGGAGAAATCCAAATGATGATATATPAGAGGCTATTTCACAG 417

Qy 386 ACGGTACCCAGAAAAACAGCAGTTGTAGTCAGAAAAAGTCGCATATG 433
Db 416 AGGATTATCATGTACATGATGAGCAGAGCAGAAACAGAAAAAGCTGG 369

RESULT 7
US-10-767-471-10718/c
Sequence 10718, Application US/10767471
GENERAL INFORMATION:

APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CL001505
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10718
LENGTH: 76573
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(76573)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10718

Query Match 7.1%; Score 32.4; DB 6; Length 76573;
Best Local Similarity 55.6%; Pred. No. 4.3;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

Qy 326 ACGGTAATAGTGCAGCTATTATCCAGAAAGTTCTGAAATPAGGCCCAATTATCCCACT 385
Db 69821 ATGGAGATGAGGCGGAGAAATCCAAATGATGATATATPAGAGGCTATTTCACAG 69762
Qy 386 ACGGTACCCAGAAAAACAGCAGTTGTAGTCAGAAAAAGTCGCATATG 433
Db 69761 AGGATTATCATGTACATGATGAGCAGAGCAGAAACAGAAAAAGCTGG 69714

RESULT 8
US-10-767-795-2887
Sequence 2887, Application US/10767795
GENERAL INFORMATION:

APPLICANT: KOVALLIC, David K.
APPLICANT: Cao, Yongwei
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(53534)B
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 117596
SEQ ID NO 2887
LENGTH: 1310
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: clone ID: GOSHI-09MAY01-C71322_1
US-10-767-795-2887

Query Match 7.0%; Score 32; DB 6; Length 1310;
Best Local Similarity 55.4%; Pred. No. 1.8;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2 TGAAGAAACAATTGTATTATATGATGTTGACAAATATCTGGTCCCTGGGATTGCCAACC 61
Db 626 TGATGATCATCTTCTCCCAATGAGATGGAGACATGCTCGAAGAAATGCTATGATGCG 685

Qy 62 CGACAAATATGATCTGCTGCTGTCAGAAATTAATTTGGCGTAATGAAATT 113
Db 686 TGTCAAAATTCAGACCTGCCCAACTCAACGATTAACAATGCTCAAAATGAAGT 737

RESULT 9
US-10-775-169-342/c
Sequence 342, Application US/10775169
GENERAL INFORMATION:

APPLICANT: WYETH
APPLICANT: Burczynski, Michael
APPLICANT: Twine, Natalie
APPLICANT: Dörner, Andrew
TITLE OF INVENTION: Method for Monitoring Drug Activities in Vivo
FILE REFERENCE: AM101080 (031896-013000)
CURRENT FILING DATE: 2004-02-11
NUMBER OF SEQ ID NOS: 5278
SOFTWARE: PatentIn version 3.2
SEQ ID NO 342
LENGTH: 5495
TYPE: DNA
ORGANISM: Homo sapiens
US-10-775-169-342

Query Match 7.0%; Score 31.8; DB 6; Length 5495;
Best Local Similarity 51.0%; Pred. No. 3.1;
Matches 75; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 2 TGAAGAAACAATTGTATTATATGATGTTGACAAATATCTGGTCCCTGGGATTGCCAACC 61
Db 4776 TGCAAAACAGAGATTTCTCGTTTGAAAAAACAAGCAAGAAACAGAGTATATCA 4717
Qy 62 CGACAAATATGATCTGCTGCTGTCAGAAATTAATTTGGCGTAATGAAATTAAGCAAGT 121

Db 4716 CGTAACGCTTGTATGAGGAAAATATGACCTTAAGTATGCTGTAATGATTTGAAAAG 4657
Qy 122 CTTCATTTAATCAGCGCCCATTTATG 148
Db 4656 CTCCTTTGATGATGAGCGGCTTTTGTG 4630

RESULT 10

US-10-767-471-8837/c
; Sequence 8837, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8837
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471-8837

Query Match
Best Local Similarity 6.8%; Score 31.2; DB 6; Length 201;
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ACGGTAATAGTCAGCTTATTCAGAAAGTCTTGAAATTAAGGCAATATTAACCACT 385
Db 183 ATGGAGATATGAGAGCCAGAAATCCAGATGTGATATATGAGAGGCTATTTCCAGG 124
Qy 386 ACGGTAACCAAAAACAGCAGTTGTAGTCAGAAACAGTCGCATATG 433
Db 123 AGGATTTTCATGTATGATGATGAGCAGACAGCAAGAAAAGCTGG 76

RESULT 11

US-10-767-471-34439/c
; Sequence 34439, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34439
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471-34439

Query Match
Best Local Similarity 6.8%; Score 31.2; DB 6; Length 201;
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ACGGTAATAGTCAGCTTATTCAGAAAGTCTTGAAATTAAGGCAATATTAACCACT 385
Db 183 ATGGAGATATGAGAGCCAGAAATCCAGATGTGATATATGAGAGGCTATTTCCAGG 124
Qy 386 ACGGTAACCAAAAACAGCAGTTGTAGTCAGAAACAGTCGCATATG 433
Db 123 AGGATTTTCATGTATGATGATGAGCAGACAGCAAGAAAAGCTGG 76

RESULT 12

US-60-548-091-5725
; Sequence 5725, Application US/60548091

; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001506
; CURRENT APPLICATION NUMBER: US/60/548,091
; NUMBER OF SEQ ID NOS: 24433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5725
; LENGTH: 394468
; TYPE: DNA
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; LOCATION: (1) -(394468)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-;
US-60-548-091-5725

Query Match
Best Local Similarity 6.8%; Score 31; DB 7; Length 394468;
Matches 55; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 233 ATGGGCGAAGTCCAGCCAGGAGGAATTAATTAATTTGCTATATTGACCAACGGCA 292
Db 320889 ATCTCTTGATAGTGAAGAGGCTGTGATATACGCTTTGCTACACAGGTCAACACAGCTG 320948
Qy 293 ATGCCAAGCATCCAGTATATGCAAAAGCGCTTAC 327
Db 320949 GTGTCAATTAACCCAGTAAAGTCCCAAGTCTTTC 320983

RESULT 13

US-10-257-047-43
; Sequence 43, Application US/10257047
; GENERAL INFORMATION:
; APPLICANT: STEWARD, MICHAEL
; TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
; FILE REFERENCE: 37945-0041
; CURRENT APPLICATION NUMBER: US/10/257,047
; PRIOR FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/GB01/01599
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: GB 0008582.9
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-10-257-047-43

Query Match
Best Local Similarity 6.8%; Score 30.8; DB 6; Length 309;
Matches 65; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 2 TGAAGAAAGAAATTTATTTATGATGTTGACAACTAGTGAGGCTGGATTGCAACG 61
Db 132 TGAAGAAAGAAAGCTTCTGTTAAGCAAGAAACAGGTGATCGACGCTTAAGTG 191
Qy 62 GCAGAAATTAATGATCTGCTGCTGTTGAGAAATTAATTTCCGTTAATTAAGCAAGT 121
Db 192 CACCGAAGAAAGACTGTGTTCTTAACGAAAGAAAGATTAATCTTCCGATGTAAGCCAGA 251
Qy 122 CT 123
Db 252 CT 253

RESULT 14

US-10-257-047-46

; Sequence 46, Application US/10257047
; GENERAL INFORMATION:
; APPLICANT: STEWARD, MICHAEL
; APPLICANT: COX, VIVIANNE FRANCES
; TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
; FILE REFERENCE: 37945-0041
; CURRENT APPLICATION NUMBER: US/10/257,047
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/GB01/01599
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: GB 0008582.9
; PRIOR FILING DATE: 2000-04-08
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 46
; LENGTH: 309
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-10-257-047-46

Query Match 6.8%; Score 30.8; DB 6; Length 309;

Best Local Similarity 53.3%; Pred.No. 2.7;
Matches 65; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 2 TGAACAAATTTGTTATTTATGATGTGACATATCTGGGTGCGCTGGGATTCACCG 61
Db 132 TGAGAACCCAAACCTACTCTGTAACGAGAACACGCTGATCGACGCTAAGTG 191
Qy 62 CGACAATTATGATCTGCTGCTTGAATATATTTGCGTAATGATTAAGCAAGT 121
Db 192 CACCGAAGAGACTGTGTTCTAAGGAAAGATTAATTGCGAATGTAAGCCAGA 251
Qy 122 CT 123
Db 252 CT 253

RESULT 15

US-10-257-047-44
; Sequence 44, Application US/10257047
; GENERAL INFORMATION:
; APPLICANT: STEWARD, MICHAEL
; APPLICANT: COX, VIVIANNE FRANCES
; TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
; FILE REFERENCE: 37945-0041
; CURRENT APPLICATION NUMBER: US/10/257,047
; CURRENT FILING DATE: 2002-10-08
; PRIOR APPLICATION NUMBER: PCT/GB01/01599
; PRIOR FILING DATE: 2001-04-09
; PRIOR APPLICATION NUMBER: GB 0008582.9
; PRIOR FILING DATE: 2000-04-08
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 3147
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Vector
; OTHER INFORMATION: PVK104-01
US-10-257-047-44

Query Match 6.8%; Score 30.8; DB 6; Length 3147;

Best Local Similarity 53.3%; Pred.No. 5.3;
Matches 65; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 2 TGAACAAATTTGTTATTTATGATGTGACATATCTGGGTGCGCTGGGATTCACCG 61
Db 198 TGAGAACCCAAACCTACTCTGTAACGAGAACACGCTGATCGACGCTAAGTG 257
Qy 62 CGACAATTATGATCTGCTGCTTGAATATATTTGCGTAATGATTAAGCAAGT 121
Db 258 CACCGAAGAGACTGTGTTCTAAGGAAAGATTAATTGCGAATGTAAGCCAGA 317

Qy 122 CT 123
Db 318 CT 319

Search completed: March 17, 2004, 08:25:32
Job time : 71.6301 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-3

Perfect score: 456

Sequence: 1 atgaactttaaaagtaga.....cgacgcctcacagactactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_NA_Main:*

```
1: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US096_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.*
```

```
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq.*
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

	Score	Match Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-3
2	453	99.3	453	92	Sequence 3, Appl
3	453	99.3	453	101	Sequence 1, Appl
4	452.8	99.3	452	13	US-60-444-371-1
5	452.8	99.3	675	13	US-08-978-878-3
6	452.8	99.3	675	31	Sequence 3, Appl
7	360	86.8	396	13	US-08-978-878-3
8	276.8	60.7	360	31	US-09-741-878-1
9	276.8	60.7	477	17	US-09-552-691-172
10	276.8	60.7	477	17	US-09-552-691-172
11	254.4	55.8	456	6	US-10-417-886-172
12	254.4	55.8	456	6	US-08-233-642A-56
13	238.4	52.3	456	23	US-09-543-407-1
14	236.8	51.9	456	23	US-09-543-407-23
15	230.4	50.5	456	23	US-09-543-407-19
16	230.4	50.5	456	23	US-09-543-407-17
17	228.8	50.2	456	23	US-09-543-407-21
18	227.2	49.8	456	23	US-09-543-407-29
19	225.6	49.8	456	23	US-09-543-407-15
20	222.4	48.8	456	23	US-09-543-407-15
21	220.8	48.4	456	23	US-09-543-407-25
22	216	47.4	456	23	US-09-543-407-11
23	173.4	38.0	361	6	US-08-233-642A-54
24	100	21.9	100	45	US-10-146-928-78
25	100	21.9	100	45	US-10-146-928-77
26	100	21.9	100	45	US-10-146-928-79
27	81.8	17.9	522	17	US-09-252-691-170
28	81.8	17.9	522	17	US-09-252-691-170
29	81.8	17.9	522	51	US-10-417-886-170
30	43.2	9.5	1083	1	PCT-US01-08631-12649
31	42.6	9.5	37551	81	US-60-245-201-6
32	42.6	9.3	37551	82	US-60-258-252-2
33	42.6	9.3	135259	1	PCT-US02-25766-8350
34	42.6	9.3	135259	47	US-10-240-442-1585
35	42.6	9.3	135259	99	US-60-423-586-24
36	42.6	9.3	135259	99	US-60-423-586-24
37	41.6	9.1	561	17	US-09-252-691-171
38	41.6	9.1	561	17	US-09-252-691-171
39	41.6	9.1	561	51	US-10-417-886-171
40	39	8.6	1360	48	US-10-017-161-1781
41	39	8.6	1360	48	US-10-292-798-417
42	39	8.6	2730	1	PCT-US01-08631-23586
43	38.2	8.4	625	43	US-10-029-386-24675
44	38.2	8.4	32768	77	US-60-213-177-372
45	38.2	8.4	87681	82	US-60-258-2175-239

RESULT 1
US-09-543-407-3

```

Sequence 3, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 456
TYPE: DNA
ORGANISM: E. Coli

```

Query Match

Best Local Similarity	100.0%;	score 456;	DB 23;	length 456;	
Matches	456;	Conservative	0;	Mismatches	0;

[illegible]

db
1 ATCTAAGCTCTGCA 60

.....GCHGAGGAGTACCGCTCTGCA 60

12

...GCGTACCGGTGGTGGCGTATATAGCGC 12

|||||||CAGGCTGGCGGTACTCTGCACCTTGCCTCTGCAA 180

221 CCAATCTCAGCTGACATTTACCACTACGGTGGCGTACTCTGCATTGCTTCC 100

181 ACUGATGCCGTA CTCTGACTTGACTATTACCCAGCA TGCGGGCTA TCCGCGC

181 ACTGATGCCCTACTCTGACTTGACTTATACCCAGATCCCCCCTTTT

241 GTTGGTCAGGCGCTCAGATGACAGCTTAATCCATTCTTTA

241 GTTGGTCAGGCTCAGATGACACCTCAATTGCATTTCC

[illegible]

b

301 GCTACCTCCTCCATCAAGTTCGGTGGT 360

VGACGGTAAACAGTTCCGTGGT 360

353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

...GCACTCACTCCGTCACGTGACTCAGGTT 420

456 CAGTACTAA

... 550...1861AACGCGCTCATCAGTACTAA 456

RESULT 2

Sequence 1 Annotated

GENERAL INFORMATION:
APPLICANT: 1131

APPLICANT: Matt, Chapman

FILE REFERENCE: WSHU 2059

CURRENT FILING DATE: 2002-01-30

SOFTWARE: PatentIn version 3.3

LENGTH: 453

TYPE: DNA
ORGANISM: E. coli

60-352-946-1

Query Match 99.3% Score 453 20 20

atches 453; Conservativ 0
 100.0%; pred. No. 7.5e-129;

	0;
1 ATGGAACCTTTTAAAGCGAATGG	Gaps 0;

1 ATCTGTTTCTGGTATCCCGTAGCGCTTGCA 60

.....GAGATCTCCGGTAGCGGCTTGGA 60

.....GCGGACCGGTCGCGGTAATAAGCGC 120

 ...GCGGTACCAAGTGGCGTAATATAGCGC 120

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

421 CCAAAATTCAGCTGAACATTACCGTACGGTGGCGTAACTCTGCACCTTGTCTCTCA 180

QY 181 ACTGATGCCCGGCTCACTGATGCTGATTAATCCAGCATGGGCGGGTAAATGATGACAT 240
DB 181 ACTGATGCCCGGCTCACTGATGCTGATTAATCCAGCATGGGCGGGTAAATGATGACAT 240
QY 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCAACGTGGCTTCGGTAAACAGC 300
DB 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCAACGTGGCTTCGGTAAACAGC 300
QY 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGGT 360
DB 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGGT 360
QY 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTACTCCGTGCAAGCTGACAGTT 420
DB 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTACTCCGTGCAAGCTGACAGTT 420
QY 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453
DB 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453

RESULT 3

US-60-444-371-1
; Sequence 1, Application US/60444371
; GENERAL INFORMATION:
; APPLICANT: Hultgren, Scott J
; TITLE OF INVENTION: BACTERIAL MODEL SYSTEM FOR AMYLOID FORMATION
; FILE REFERENCE: WSHU 2059.1
; CURRENT APPLICATION NUMBER: US/60/444,371
; CURRENT FILING DATE: 2003-01-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Escherichia coli
US-60-444-371-1

Query Match 99.3%; Score 453; DB 101; Length 453;
Best Local Similarity 100.0%; Pred. No. 7.5e-129;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTTCTCCGGTAGCGCTCTGGCA 60
DB 1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTTCTCCGGTAGCGCTCTGGCA 60
QY 61 GGTGTTGTTCTCAGTACGGCGCGGGGTAAACACAGGTGGTGGCGGTAAATATAGCGGC 120
DB 61 GGTGTTGTTCTCAGTACGGCGCGGGGTAAACACAGGTGGTGGCGGTAAATATAGCGGC 120
QY 121 CCAAAATTCGAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGCTGCA 180
DB 121 CCAAAATTCGAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGCTGCA 180
QY 181 ACTGATGCCCGGCTCACTGACTGACTTAATCCAGCAATGGGCGGGTAAATGATGACAT 240
DB 181 ACTGATGCCCGGCTCACTGACTGACTTAATCCAGCAATGGGCGGGTAAATGATGACAT 240
QY 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCAACGTGGCTTCGGTAAACAGC 300
DB 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCAACGTGGCTTCGGTAAACAGC 300
QY 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGGT 360
DB 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGGT 360
QY 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTACTCCGTGCAAGCTGACAGTT 420
DB 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTACTCCGTGCAAGCTGACAGTT 420
QY 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453
DB 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453

DB 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453

RESULT 4

US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8601723-1
; EARLIER FILING DATE: 1988-05-06
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-08-978-878-3

Query Match 99.3%; Score 452.8; DB 13; Length 675;
Best Local Similarity 99.6%; Pred. No. 9.9e-129;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTTCTCCGGTAGCGCTCTGGCA 60
DB 83 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTTCTCCGGTAGCGCTCTGGCA 142
QY 61 GGTGTTGTTCTCAGTACGGCGCGGGGTAAACACAGGTGGTGGCGGTAAATATAGCGGC 120
DB 143 GGTGTTGTTCTCAGTACGGCGCGGGGTAAACACAGGTGGTGGCGGTAAATATAGCGGC 202
QY 121 CCAAAATTCGAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGCTGCA 180
DB 203 CCAAAATTCGAGCTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGCTGCA 262
QY 181 ACTGATGCCCGGCTCACTGACTGACTTAATCCAGCAATGGGCGGGTAAATGATGACAT 240
DB 263 ACTGATGCCCGGCTCACTGACTGACTTAATCCAGCAATGGGCGGGTAAATGATGACAT 322
QY 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCAACGTGGCTTCGGTAAACAGC 300
DB 323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCAACGTGGCTTCGGTAAACAGC 382
QY 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGGT 360
DB 383 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGGT 442
QY 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTACTCCGTGCAAGCTGACAGTT 420
DB 443 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTACTCCGTGCAAGCTGACAGTT 502
QY 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 456
DB 503 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 538

RESULT 5

US-09-741-873B-3
; Sequence 3, Application US/09741873B
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873B
; PRIOR APPLICATION NUMBER: 2003-04-04
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873B-3

Query Match

Best Local Similarity 99.3%; Score 452.8; DB 31; Length 675;
Matches 454; Conservative 99.6%; Pred. No. 9,9e-129;

1 ATGAACCTTTAAAGTAAAGCAATTTGAGCAATCTCTCGTAGCGCTGCA 60
83 ATGAACCTTTAAAGTAAAGCAATTTGAGCAATCTCTCGTAGCGCTGCA 142
61 GGTGTTCTCTAGTACGCGCGCGGTAACCGGTTGCGGTAATTAATAGCGC 120
143 GGTGTTCTCTAGTACGCGCGCGGTAACCGGTTGCGGTAATTAATAGCGC 202
121 CCAAAATTCGAGCTGAACATTTACAGTACGCTGCGGTAATCTCTGCA 180
203 CCAAAATTCGAGCTGAACATTTACAGTACGCTGCGGTAATCTCTGCA 262
181 ACTGATGCCCGTAACTGACTGATTAACAGTACGCTGCGGTAATCTCTGCA 240
263 ACTGATGCCCGTAACTGACTGATTAACAGTACGCTGCGGTAATCTCTGCA 322
241 GTTGTGAGGCTCAGATGACGCTCAATGATCTGACCAACGCTGCTCGTAACAGC 300
323 GTTGTGAGGCTCAGATGACGCTCAATGATCTGACCAACGCTGCTCGTAACAGC 382
301 GCTACTCTTATCATGTTGAAACGCGCAAAATTCGAAATGACGTTGCTGCTGCT 360
383 GCTACTCTTATCATGTTGAAACGCGCAAAATTCGAAATGACGTTGCTGCTGCT 442
361 GCGAACGCTGCTCAGTTGACGCTGATCTGATCTGCTGCTGCTGCTGCTGCT 420
443 GCGAACGCTGCTCAGTTGACGCTGATCTGATCTGCTGCTGCTGCTGCTGCT 502
421 GCGTTTGTATCAACGCGACCGCTCATCTAGTACTAA 456
503 GCGTTTGTATCAACGCGACCGCTCATCTAGTACTAA 538

RESULT 6

US-08-978-878-1
; Sequence 1, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; PRIOR APPLICATION NUMBER: 1997-11-26
; PRIOR FILING DATE: 1988-05-06
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; PRIOR APPLICATION NUMBER: US 08/495,959
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-08-978-878-1

Query Match
Best Local Similarity 86.8%; Score 396; DB 13; Length 396;
Matches 396; Conservative 100.0%; Pred. No. 3e-111;

61 GGTGTTCTCTAGTACGCGCGCGGTAACCGGTTGCGGTAATTAATAGCGC 120
1 GGTGTTCTCTAGTACGCGCGCGGTAACCGGTTGCGGTAATTAATAGCGC 60
121 CCAAAATTCGAGCTGAACATTTACAGTACGCTGCGGTAATCTCTGCA 180
61 CCAAAATTCGAGCTGAACATTTACAGTACGCTGCGGTAATCTCTGCA 120
181 ACTGATGCCCGTAACTGACTGATTAACAGTACGCTGCGGTAATCTCTGCA 240
121 ACTGATGCCCGTAACTGACTGATTAACAGTACGCTGCGGTAATCTCTGCA 180
241 GTTGTGAGGCTCAGATGACGCTCAATGATCTGACCAACGCTGCTCGTAACAGC 300
301 GCTACTCTTATCATGTTGAAACGCGCAAAATTCGAAATGACGTTGCTGCTGCT 360
241 GCTACTCTTATCATGTTGAAACGCGCAAAATTCGAAATGACGTTGCTGCTGCT 300
361 GCGAACGCTGCTCAGTTGACGCTGATCTGATCTGCTGCTGCTGCTGCTGCT 420
301 GCGAACGCTGCTCAGTTGACGCTGATCTGATCTGCTGCTGCTGCTGCTGCT 420
421 GCGTTTGTATCAACGCGACCGCTCATCTAGTACTAA 456
361 GCGTTTGTATCAACGCGACCGCTCATCTAGTACTAA 396

RESULT 7

US-09-741-873B-1
; Sequence 1, Application US/09741873B
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

FILE REFERENCE: 012889-084
 CURRENT APPLICATION NUMBER: US/09/741,873B
 PRIOR FILING DATE: 2003-04-04
 PRIOR APPLICATION NUMBER: SE 8801723-1
 PRIOR FILING DATE: 1998-05-06
 PRIOR APPLICATION NUMBER: US 08/978,878
 PRIOR FILING DATE: 1997-11-26
 PRIOR APPLICATION NUMBER: US 07/347,189
 PRIOR FILING DATE: 1988-05-04
 PRIOR APPLICATION NUMBER: US 07/789,437
 PRIOR FILING DATE: 1991-11-06
 PRIOR APPLICATION NUMBER: US 07/970,846
 PRIOR FILING DATE: 1992-11-03
 PRIOR APPLICATION NUMBER: US 08/187,865
 PRIOR FILING DATE: 1994-01-28
 PRIOR APPLICATION NUMBER: US 08/318,519
 PRIOR FILING DATE: 1994-10-05
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO: 1
 LENGTH: 360
 TYPE: DNA
 ORGANISM: Escherichia coli
 US-09-741-873B-1

Query Match 78.9%; Score 360; DB 31; Length 360;
 Best Local Similarity 100.0%; Pred. No. 4e-100;
 Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 GGTGTTGTTCTCTGAGTACGGCGCGCGGTAAACACGGGTGGCGGTAAATATAGCGGC 120
 DB 1 GGTGTTGTTCTCTGAGTACGGCGCGCGGTAAACACGGGTGGCGGTAAATATAGCGGC 60
 QY 121 CCAAAATCTGAGCTGAACATTTACAGTACGGTGGCGGTAACTGCACTTGCTGCA 180
 DB 61 CCAAAATCTGAGCTGAACATTTACAGTACGGTGGCGGTAACTGCACTTGCTGCA 120
 QY 181 ACTATGCGCCGTAATCTGACTGATCTATTAACCGACATGGCGGTAAATGTCACAT 240
 DB 121 ACTATGCGCCGTAATCTGACTGATCTATTAACCGACATGGCGGTAAATGTCACAT 180
 QY 241 GTTGGTCAGGGCTCAGATGACGCTCATGCTGACCCAACTGGCTGGTAAACAGC 300
 DB 181 GTTGGTCAGGGCTCAGATGACGCTCATGCTGACCCAACTGGCTGGTAAACAGC 240
 QY 301 GCTACTCTTGTACGTGAGCAACGCAAAATCTGAAATGACGGTTAAACAGTTGGTGT 360
 DB 241 GCTACTCTTGTACGTGAGCAACGCAAAATCTGAAATGACGGTTAAACAGTTGGTGT 300
 QY 361 GGCAACGGTGTGCGAGTTGACCACTGACATCTCTCCGTTAACTGACTAGATT 420
 DB 301 GGCAACGGTGTGCGAGTTGACCACTGACATCTCTCCGTTAACTGACTAGATT 360

RESULT 8

US-09-252-691-172

Sequence 172; Application US/09252691B

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstock et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER

FILE REFERENCE: 107196.135

CURRENT APPLICATION NUMBER: US/09/252,691B

PRIOR FILING DATE: 1999-02-18

NUMBER OF SEQ ID NOS: 11324

SEQ ID NO 172

LENGTH: 477

TYPE: DNA

ORGANISM: Enterobacter cloacae

Query Match 60.7%; Score 276.8; DB 17; Length 477;
 Best Local Similarity 75.4%; Pred. No. 2.4e-74;

Matches 344; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTAGAACAATTGACCAATGTAATCTCGGTAGCGCTTGCA 60
 DB 22 ATGAACCTTTTCAAGTAGGAGTATATGACCAATGTAATCTCGGTAGCGCTTGCA 81
 QY 61 GGTGTTGTTCTCTGAGTACGGCGCGCGGTAAACACGGGTGGCGGTAAATATAGCGGC 120
 DB 82 GGTGTTGTTCTCTGAGTACGGCGCGCGGTAAACACGGGTGGCGGTAAATATAGCGGC 141
 QY 121 CCAAAATCTGAGCTGAACATTTACAGTACGGTGGCGGTAACTGCACTTGCTGCA 180
 DB 142 CCAAAATCTGAGCTGAACATTTACAGTACGGTGGCGGTAACTGCACTTGCTGCA 201
 QY 181 ACTATGCGCCGTAATCTGACTGATCTATTAACCGACATGGCGGTAAATGTCACAT 240
 DB 202 ACTATGCGCCGTAATCTGACTGATCTATTAACCGACATGGCGGTAAATGTCACAT 261
 QY 241 GTTGGTCAGGGCTCAGATGACGCTCATGCTGACCCAACTGGCTGGTAAACAGC 300
 DB 262 GTTGGTCAGGGCTCAGATGACGCTCATGCTGACCCAACTGGCTGGTAAACAGC 321
 QY 301 GCTACTCTTGTACGTGAGCAACGCAAAATCTGAAATGACGGTTAAACAGTTGGTGT 360
 DB 322 GCTACTCTTGTACGTGAGCAACGCAAAATCTGAAATGACGGTTAAACAGTTGGTGT 381
 QY 361 GGCAACGGTGTGCGAGTTGACCACTGACATCTCTCCGTTAACTGACTAGATT 420
 DB 382 GGCAACGGTGTGCGAGTTGACCACTGACATCTCTCCGTTAACTGACTAGATT 441
 QY 421 GCTTTGGTAAACAGCGACCGCTCATGAGTAACTA 456
 DB 442 GCTTTGGTAAACAGCGACCGCTCATGAGTAACTA 477

RESULT 9

US-09-252-691C-172

Sequence 172; Application US/09252691C

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstock et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER

FILE REFERENCE: 107196.135

CURRENT APPLICATION NUMBER: US/09/252,691C

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/094,145

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 60/074,787

NUMBER OF SEQ ID NOS: 11326

SEQ ID NO 172

LENGTH: 477

TYPE: DNA

ORGANISM: Enterobacter cloacae

US-09-252-691C-172

Query Match 60.7%; Score 276.8; DB 17; Length 477;

Best Local Similarity 75.4%; Pred. No. 2.4e-74;

Matches 344; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTAGAACAATTGACCAATGTAATCTCGGTAGCGCTTGCA 60
 DB 22 ATGAACCTTTTCAAGTAGGAGTATATGACCAATGTAATCTCGGTAGCGCTTGCA 81
 QY 61 GGTGTTGTTCTCTGAGTACGGCGCGCGGTAAACACGGGTGGCGGTAAATATAGCGGC 120
 DB 82 GGTGTTGTTCTCTGAGTACGGCGCGCGGTAAACACGGGTGGCGGTAAATATAGCGGC 141
 QY 121 CCAAAATCTGAGCTGAACATTTACAGTACGGTGGCGGTAACTGCACTTGCTGCA 180
 DB 142 CCAAAATCTGAGCTGAACATTTACAGTACGGTGGCGGTAACTGCACTTGCTGCA 201
 QY 181 ACTATGCGCCGTAATCTGACTGATCTATTAACCGACATGGCGGTAAATGTCACAT 240

```

Db      202 AGGAGCGCCGAGATTCTTAATTGACATTACCGACATGGTGGTAAAGCGGCAAT 261
Qy      241 GTTGTGAGGCTCAGATGACAGCTCAATGATCTGACCCACGTGGCTCGGTAAACAGC 300
Db      262 GTTGGCCAGGGCTGATGATGACAGTTCTATGATCTGGTCAAAAAGGGCTTGGTAAACGC 321
Qy      301 GCTACTCTGATGATGAAACGGCAAAATTTCTGAAATGACGGTTAAACAGTTCCGTGTGT 360
Db      322 GGCACCATGACCAATGAAATGACAAAGACTCTGTTATCAACGTGAAAACAGTTCCGGCGC 381
Qy      361 GGCACCGGTGCTGACGATTGACAGCTGATCTTAATCTCCGCTCAACGTGACTGAGTT 420
Db      382 GGCACCGGGGGGGGGGTGACAGACAGACGCTCCGGCTCAACGGTACTGTGACCAAGGTT 441
Qy      421 GCGTTTGGTAAACAGCGGACCGCTCATGACTAGTAA 456
Db      442 GCGTTTGGCAACACGCGGACCGCACACAGTACTGA 477

```

RESULT 10

```

US-10-417-886-172
/ Sequence 172, Application US/10417886
/ GENERAL INFORMATION:
/ APPLICANT: Keith G. Weinstein et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
/ FILE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.135
/ CURRENT APPLICATION NUMBER: US/10/417.886
/ CURRENT FILING DATE: 2003-04-17
/ PRIOR APPLICATION NUMBER: US/09/252.691C
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,145
/ PRIOR FILING DATE: 1998-07-24
/ PRIOR APPLICATION NUMBER: US 60/074,787
/ PRIOR FILING DATE: 1998-02-18
/ NUMBER OF SEQ ID NOS: 11326
/ SEQ ID NO 172
/ LENGTH: 477
/ TYPE: DNA
/ ORGANISM: Enterobacter cloacae
US-10-417-886-172

```

```

Query Match      60.7%; Score 276.8; DB 51; Length 477;
Best Local Similarity 75.4%; Pred. No. 2.4e-74;
Matches 344; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```

```

Qy      1 ATGAACCTTTAAAGTGAAGCAATTGACGAATCGTATTCTCGGTAGCGCTTGGA 60
Db      22 ATGAACCTTTCAAGTGGAGGAGTTATGACGAATCGTATTCTCGGTAGCTTGCA 81
Qy      61 GGTGTGTTCTCTGATAGCGGCGGCGGTAAACAGGTTGGCGGTAAATTAATAGCGG 120
Db      82 GGTGCGGTACCAATTTGGGCGGCGGCGGCGGCGGTGGGCTGGCGGCAATTAACGGC 141
Qy      121 CCAAAATTTAGTGAACATTTACAGTACGTTGCGGTAACTCTGACTTGGCTGCA 180
Db      142 CTTACTACCAACCTGACCATTTACAGTACGCGGCGGTAACTCGCGCTTGGCTTGCAG 201
Qy      181 ACTGATGCCGTAACTGATCTGATCTTAACTTAACTTAACTGATGCGGCGGTAAATGTCAGAT 240
Db      202 AGGAGCGCCGAGATTTGAAATGACCAATTAACAGATGATGATGTTGTTAGCGGAGAT 261
Qy      241 GTTGTGAGGCTCAGATGACAGCTCAATGATCTGACCCACGTGGCTCGGTAAACAGC 300
Db      262 GTTGGCCAGGGCTGATGATGACAGTTCTATGATCTGGTCAAAAAGGGCTTGGTAAACGC 321
Qy      301 GCTACTCTGATGATGAAACGGCAAAATTTCTGAAATGACGGTTAAACAGTTCCGTGTGT 360
Db      322 GGCACCATGACCAATGAAATGACAAAGACTCTGTTATCAACGTGAAAACAGTTCCGGCGC 381
Qy      361 GGCACCGGTGCTGACGATTGACAGCTGATCTTAATCTCCGCTCAACGTGACTGAGTT 420

```

```

Db      382 GGCACCGCGCGGGGTAGACAGACAGCGTCCGGCTCAACGGTGACTGTGACCAAGTT 441
Qy      421 GCGTTTGGTAAACACGCGGACCGCTCATGACTGACTAA 456
Db      442 GCGTTTGGCAACACGCGGACCGCACACAGTACTGA 477

```

RESULT 11

```

US-08-233-642A-56
/ Sequence 56, Application US/08233642A
/ GENERAL INFORMATION:
/ APPLICANT: Kay, William W.
/ APPLICANT: Collinson, S. Karen
/ APPLICANT: Clouthier, Sharon C.
/ APPLICANT: Doran, James L.
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
/ TITLE OF INVENTION: BASED VACCINES
/ NUMBER OF SEQUENCES: 58
/ NUMBER OF SEQUENCES: 58
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Seed and Berry
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: U.S.A.
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/233,642A
/ FILING DATE: 26-APR-1994
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: King, Joshua
/ REGISTRATION NUMBER: 35,570
/ REFERENCE/DOCKET NUMBER: 920043.403C3
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ TELEX: 3723836 SEEDANBERY
/ INFORMATION FOR SEQ ID NO: 56:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 456 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..456
US-08-233-642A-56

```

```

Query Match      55.8%; Score 254.4; DB 6; Length 456;
Best Local Similarity 72.4%; Pred. No. 2e-67;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```

```

Qy      1 ATGAACCTTTAAAGTGAAGCAATTGACGAATCGTATTCTCGGTAGCGCTTGGA 60
Db      1 ATGAACCTTTAAAGTGGAGCATTTCCAGCAATCGTATTCTGGCAGTGTGCTGCT 60
Qy      61 GGTGTGTTCTCTGATAGCGGCGGCGGTAAACAGGTTGGCGGTAAATTAATAGCGG 120
Db      61 GCGTGTGTTCAACATGGGCGGCGGCGGTATCATTAACGCGGCGGCAATGATTTCCGCGC 120
Qy      121 CCAAAATTTAGTGAACATTTACAGTACGTTGCGGTAACTCTGACTTGGCTGCA 180
Db      121 CCGACTCAAGTGTGAGCATTTATCAGTACGTTCCGCTTAACGCTGCCCTTGTCTGCA 180
Qy      181 ACTGATGCCGTAACTGATCTGATCTTAACTTAACTTAACTGATGCGGCGGTAAATGTCAGAT 240
Db      181 AGCGATGCCGTAAATCTGAAACGACCATTAACCAAGCGGCTTATGTGAACGGGCGCGAT 240

```

```

; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.

```


;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543.407
;; CURRENT FILING DATE: 2000-04-05
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 19
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
;; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-19

Query Match 51.9%; Score 236.8; DB 23; Length 456;
Best Local Similarity 70.0%; Pred. No. 5.4e-62;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

```
QY 1 ATGAACCTTTTAAAGTAGAACAATTCAGCAATCGTATTCCTCCGTAAGCGCTTGCA 60
Db 1 ATGAACCTTTTAAAGTAGAACAATTCAGCAATCGTATTCCTCCGTAAGCGCTTGCA 60
QY 61 GGTGTCTTCTCAGTACGGGCGGCGGTACCAAGCGTGTGGCGGTAAATAGCGGC 120
Db 61 GGTGTCTTCTCAGTACGGGCGGCGGTAAATAGTAAACGCGGCAATAGTTCGGC 120
QY 121 CCAAACTTCTGAGTGAACATTTACAGTACGCTGGCGGTAACTCTGCACTTGTCTGCA 180
Db 121 CCGGACTCAACGTGAGCATTTTACAGTACGCTGGCGGTAACTCTGCACTTGTCTGCA 180
QY 181 ACTGATGCCGTAATCTGACTGATTAATCCAGATGCGCGGTAAATGTCAGAT 240
Db 181 AGCGATGCCGTAATGATGATCACTGATTAATCCGTTGTTACCTGAAATGTCAGAT 240
QY 241 GTTGTGAGGGCTCAGTACAGTCAATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 GCAAGCCAGGGTGGGTAATATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCTACTCTTATCAGTGAAGCAAGCAAAATTTCTGAATGACGGTTAAACAGTGTGT 360
Db 301 GCAACCATGACAGTGAAGCAAGCAAAATTTCTGAATGACGGTTAAACAGTGTGT 360
QY 361 GGCACGCGTCTGAGTGAACAGTCACTTAATCTCTCCGTCACAGTCACTGAGTT 420
Db 361 AATTAACCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTGCGTCAAGTT 420
QY 421 GCGTTGGTAAACAAAGCGACCGCTCATCAGACTAA 456
Db 421 GGTTTGGCAACAAAGCGACCGCTCATCAGACTAA 456
```

RESULT 15
US-09-543-407-17

;; Sequence 17, Application US/09543407
;; GENERAL INFORMATION:
;; APPLICANT: White, Aaron P.
;; APPLICANT: Doran, James L.
;; APPLICANT: Collinson, S. Karen
;; APPLICANT: Kay, William W.
;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543.407
;; CURRENT FILING DATE: 2000-04-05
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 17
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence

;; FEATURE:
;; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
;; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-17

Query Match 50.5%; Score 230.4; DB 23; Length 456;
Best Local Similarity 69.1%; Pred. No. 5.2e-60;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
QY 1 ATGAACCTTTTAAAGTAGAACAATTCAGCAATCGTATTCCTCCGTAAGCGCTTGCA 60
Db 1 ATGAACCTTTTAAAGTAGAACAATTCAGCAATCGTATTCCTCCGTAAGCGCTTGCA 60
QY 61 GGTGTCTTCTCAGTACGGGCGGCGGTACCAAGCGTGTGGCGGTAAATAGCGGC 120
Db 61 GGTGTCTTCTCAGTACGGGCGGCGGTAAATAGTAAACGCGGCAATAGTTCGGC 120
QY 121 CCAAACTTCTGAGTGAACATTTACAGTACGCTGGCGGTAACTCTGCACTTGTCTGCA 180
Db 121 CCGGACTCAACGTGAGCATTTTACAGTACGCTGGCGGTAACTCTGCACTTGTCTGCA 180
QY 181 ACTGATGCCGTAATCTGACTGATTAATCCAGATGCGCGGTAAATGTCAGAT 240
Db 181 AGCGATGCCGTAATGATGATCACTGATTAATCCGTTGTTACCTGAAATGTCAGAT 240
QY 241 GTTGTGAGGGCTCAGTACAGTCAATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 GCAAGCCAGGGTGGGTAATATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCTACTCTTATCAGTGAAGCAAGCAAAATTTCTGAATGACGGTTAAACAGTGTGT 360
Db 301 GCAACCATGACAGTGAAGCAAGCAAAATTTCTGAATGACGGTTAAACAGTGTGT 360
QY 361 GGCACGCGTCTGAGTGAACAGTCACTTAATCTCTCCGTCACAGTCACTGAGTT 420
Db 361 AATTAACCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTGCGTCAAGTT 420
QY 421 GCGTTGGTAAACAAAGCGACCGCTCATCAGACTAA 456
Db 421 GGTTTGGCAACAAAGCGACCGCTCATCAGACTAA 456
```

Search completed: March 18, 2004, 02:46:48
Job time: 2925.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-3

Perfect score: 456

Sequence: 1 atgaacttttaaaagtaga.....cgaccgcctcactgactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Pending Patente NA New:*

1: /cgn2_6/ptodata/2/pna/BCT_NEM_COMB_seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEM_COMB_seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEM_COMB_seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEM_COMB_seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEM_COMB_seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEM_COMB_seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEM_COMB_seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	452.8	99.3	675	US-09-741-873C-3	Sequence 3, Appli
2	396	86.8	396	US-09-741-873C-1	Sequence 58, Appli
3	34.2	7.5	79977	US-10-765-790-58	Sequence 1, Appli
4	34.2	7.5	186449	US-10-021-698A-115	Sequence 715, App
5	33	7.2	439	US-10-021-698A-106	Sequence 1006, App
6	32	7.0	1355	US-10-045-674A-593	Sequence 593, App
7	30.8	6.8	290892	US-10-767-471-10809	Sequence 10809, App
8	30.4	6.7	945	US-10-767-701-7789	Sequence 5635, App
9	30.2	6.6	63693	US-60-548-091-5635	Sequence 7789, App
10	30.2	6.6	290892	US-10-767-471-10809	Sequence 10809, App
11	30	6.6	372	US-10-021-698A-1217	Sequence 1217, App
12	30	6.6	992	US-10-021-698A-796	Sequence 796, App
13	30	6.6	1203	US-10-021-698A-796	Sequence 1203, App
14	29.6	6.5	1203	US-10-021-698A-796	Sequence 1203, App
15	29.2	6.4	306	US-10-100-683-13342	Sequence 13342, App
16	29.2	6.4	306	US-10-100-683-13342	Sequence 13342, App
17	29.2	6.4	851	US-10-100-683-13343	Sequence 13343, App
18	29.2	6.4	2776	US-10-100-683-13343	Sequence 13343, App
19	29.2	6.4	4660	US-10-100-683-13343	Sequence 13343, App
20	29	6.4	507	US-10-767-795-1598	Sequence 80, Appli
21	29	6.4	551	US-10-767-795-1598	Sequence 118, App
22	29	6.4	753	US-10-767-701-9504	Sequence 9504, App
23	29	6.4	1350	US-10-767-701-9504	Sequence 863, App
24	28.8	6.3	6705	US-09-032-438C-5	Sequence 5, Appli
25	28.8	6.3	6705	US-10-336-215A-5	Sequence 5, Appli
26	28.8	6.3	6705	US-10-336-215A-5	Sequence 5, Appli

27	28.8	6.3	6705	US-10-340-097B-5	Sequence 5, Appli
28	28.8	6.3	6819	US-09-032-438C-2	Sequence 2, Appli
29	28.8	6.3	6819	US-10-336-215A-2	Sequence 2, Appli
30	28.8	6.3	6819	US-10-336-215A-2	Sequence 2, Appli
31	28.8	6.3	6819	US-10-340-097B-2	Sequence 12, Appli
32	28.8	6.3	7318	US-10-783-528-12	Sequence 1, Appli
33	28.8	6.3	7488	US-09-032-438C-1	Sequence 1, Appli
34	28.8	6.3	7783	US-10-336-215A-1	Sequence 1, Appli
35	28.8	6.3	7783	US-10-336-215A-1	Sequence 1, Appli
36	28.8	6.3	7783	US-10-340-097B-1	Sequence 1216, App
37	28.6	6.3	435	US-10-021-698A-1216	Sequence 1022, App
38	28.6	6.3	704	PCT-US04-05654-1022	Sequence 1022, App
39	28.6	6.3	860	PCT-US04-05654-1018	Sequence 1018, App
40	28.6	6.3	906	PCT-US04-05654-1020	Sequence 1020, App
41	28.6	6.3	74279	US-10-765-790-16	Sequence 16, Appli
42	28.4	6.2	524	US-10-767-701-22290	Sequence 22290, App
43	28.2	6.2	654	US-10-767-701-4153	Sequence 4153, App
44	28.2	6.2	1252	PCT-US04-05654-877	Sequence 877, App
45	28.2	6.2	1606	PCT-US04-05654-2525	Sequence 2525, App

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83) ..(538)
US-09-741-873C-3
Query Match 99.3%; Score 452.8; DB 5; Length 675;
Best Local Similarity 99.6%; Pred. No. 2.3e-145;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTAGAAGCAATTCGATTTCTCCGTAGCGCTTGCA 60
DB 83 ATGAACCTTTAAAGTAGAAGCAATTCGATTTCTCCGTAGCGCTTGCA 142
QY 61 GGTGTGTTCTCTGATCGGCGCGCGGTAAACGATGCGGTAAATGACGGC 120
DB 143 GGTGTGTTCTCTGATCGGCGCGCGGTAAACGATGCGGTAAATGACGGC 202
QY 121 CCAATTCGATGCAATTAACGATGCGGTGCGGTAACTGCACTTGCTGCA 180

Db	203	CCAAATTCGTAGCTGAACATTTCACAGTAGAGGGGGGTAACCTCTGCACCTTGGCTCGAA	262
OY	181	ACTGATGCCCGTAACTCTGACTTGACTATTTACCAGCATGCGCGGGATGATGTCAGAT	240
Db	263	ACTATATGCCGTAACTCTGACTTGACTATTTACCCAGCATGGGCGGGTAAATGGTCAGAT	322
OY	241	GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGATCGAACCCGACGGTGTGGTAAAGC	300
Db	323	GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGATCGAACCCGACGGTGTGGTAAAGC	382
OY	301	GCTACTCTTGATATGATGAGAACGGGCAGAAATTTCTGAATATGACGGTTAAACAGTTGATGGT	360
Db	383	GCTACTCTTGATATGATGAGAACGGGCAGAAATTTCTGAATATGACGGTTAAACAGTTGATGGT	442
OY	361	GGCAACGGTCTGCTCAGTTGACACAGCTGCATCTAACTCTCCGTCAACGTGACCTCAGATT	420
Db	443	GGCAACGGTCTGCTCAGTTGACACAGCTGCATCTAACTCTCCGTCAACGTGACCTCAGATT	502
OY	421	GGCTTTGGTAAACAACCGCAGCCGGTGAATGATCTAA	456
Db	503	GGCTTTGGTAAACAACCGCAGCCGGTGAATGATCTAA	538

```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As well As Its Preparation
; FILE REFERENCE: 012889-064
; CURRENT APPLICATION NUMBER: US/09/741.873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 06/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

```

Query Match	86.8%	Score 396	DB 5	Length 396
Best Local Similarity	100.0%	Pred. No.	5.2e-126	
Matches	396	Conservative	0	Mismatches 0; Indels 0; Gaps 0
QY	61	GGTGTGTTCCCACTAGCGCGCGCGCGGTAAACAACGGTGTGTGCGCGTAAATATATAGCGC	120	
Db	1	GGTGTGTTCTCAGTAGCGCGCGCGCGGTAAACAACGGTGTGTGCGCGTAAATATAGCGCGC	60	
QY	121	CCAAATTCGACCTGAACAATTTCCAGTAGCGGTGCGGTAACTCGACCTTGCTTGCA	180	
Db	61	CCAAATTCGAGCTGACATTAATTCAGTAGCGGTGCGGTAACTCGACCTTGCTTGCAA	120	
QY	181	ACTGATGCCCGTAATCTGACTTGACATTAATCCGACGATGCGCGCGGTAAATGATGCGCAT	240	
Db	121	ACTGATGCCCGTAATCTGACTTGACATTAATCCGACGATGCGCGCGGTAAATGATGCGCAT	180	
QY	241	GTGTGATGAGGCTCAGATGACAGCTCAATTCATCTGACCCAAAGTGTGCTTGATACAGC	300	

Db	181	GTGGTCAGGGCTCAATGATCAGCTCATCGATCCCAAGCGTTCGGTAACAGC	240
Qy	301	GCTACTCTTGATCATGTCGGAAACGGCAAAATTTCTGAATATGAACGTTAAACAGTTCCGTTGGT	360
Db	241	GCTACTCTTGATCATGTCGGAAACGGCAAAATTTCTGAATATGAACGTTAAACAGTTCCGTTGGT	300
Qy	361	GGCAACGGTGCCTGCACTTGAACCAAGCTCCTCACTTCCTTCCTGCAACGTTGACTCAGGTT	420
Db	301	GGCAACGGTGCCTGCACTTGAACCAAGCTCCTCACTTCCTTCCTGCAACGTTGACTCAGGTT	360
Qy	421	GGCTTTGTTAACACGCAACGGCTCATCAGTACTTAA	456
Db	361	GGCTTTGTTAACACGCAACGGCTCATCAGTACTTAA	396

```

RESULT 3
US-10-765-790-58/c
; Sequence 58, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 79977
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-765-790-58

```

```

Query Match          7.5%; Score 34.2; DB 6; Length 79977;
Best Local Similarity 48.2%;
Matches    96; Conservative      0; Mismatches 103; Indels      0; Gaps      0;

OY      61 GGTTGTTGTTCTCCAGTACGGCGCGCGCGGTAAACAAGTGTTGTCGGGTAAATATAGCGGC 120
Db       15147 GGTGGTATGATGGTATGATGATGATGATGATGATGAAGAAGATGGTATGGTATGGT 150888

OY      121 CCAATTTCGAGGTGAACAATTAAACCAGTAACGGTGGCGGTAACTCTGCACCTTCTGCCAA 180
Db       15087 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATC 150288

OY      181 ACTGATGCCCGTAACCTCTGACTTGACTAATTAACCAAGCATGGCGGCGCGGTAAATGGTCAGAT 240
Db       15027 GCGGATGGTGGTGGTATGGTGGTATGATGATGATGATGATGATGATGATGATGATGAT 14968

OY      241 GTTGGTCAGAGGCTCAGATG 259
Db       14967 GGTGGTGGTGGTGAAGATG 14949

RESULT 4
US-10-021-698A-715
; Sequence 715, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDEMEGH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
```

```

FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (129353)..(129452)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (144859)..(144958)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (155513)..(155612)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (166141)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (166388)..(166487)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (172617)..(172716)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (177984)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (178024)..(178123)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (180005)..(180104)
/ OTHER INFORMATION: a, t, c or g
FEATURE:
/ NAME/KEY: modified_base
/ LOCATION: (183290)..(183389)
/ OTHER INFORMATION: a, t, c or g
;
US-10-021-698A-715

Query Match      7.5%; Score 34.2; DB 6; Length 186449;
Best Local Similarity 43.9%; Pred. No. 0.88;
Matches 147; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

OY      46  GGTAGCCGCTCGCAGCGTGTTCCTCCTAGTACGGCGCGCGCGTAAACACGCTGTCGCC 105
DB      148605  GGTAGTGTGCGTGTGTGTATGATGATGATGATGATGATGATGATGATGATGAT 148666
OY      106  GGTAATATATAGCGCGCCAAATCTGAGCTGAACATTACAGTACGGTGGTAACTCT 165
DB      148665  GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148722
OY      166  GCACCTTGCTTCGCAACATGATGCCGTAACCTGACTGATTAACCAAGCATGCGCGC 225
DB      148725  GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148784
OY      226  GGTAATGATGACAGATGTTGTGTCAGGGCTCAAGTACAGCTCAATGATGATGACCCAACT 285
DB      148785  GGTAGTGTGCGTGTGTGTGTATCTGATGCGTGTGATGATGATGATGATGATGATGATGAT 148844
OY      286  GGCTTCGGTAAACAGCGCTACTCTTGATCAGTGAACGCGCAAAATCTGAAATGACGGTT 345
DB      148845  GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148900
OY      346  AAACAGTTTCGGTGTGCGCAACGTCCTGCAGTTGA 380
DB      148905  GGTAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148939

```

[illegible]

RESULT 6
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; ADDRESS:

APPLICANT: LADNER, ROBERT C.
 APPLICANT: COHEN, EDWARD H.
 APPLICANT: NASTRI, HORACIO G.
 APPLICANT: ROOKEY, KRISTIN L.
 APPLICANT: HOET, RENE
 APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
 TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
 TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
 TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
 FILE REFERENCE: DYA/002 CIP2
 CURRENT APPLICATION NUMBER: US/10/045,674A
 CURRENT FILING DATE: 2001-10-25
 PRIOR APPLICATION NUMBER: 06/198,069
 PRIOR FILING DATE: 2000-04-17
 PRIOR APPLICATION NUMBER: 09/637,306
 PRIOR FILING DATE: 2001-04-17
 NUMBER OF SEQ ID NOS: 635
 SOFTWARE: Patentin Ver. 2.1
 SEQ ID NO 593
 LENGTH: 1355
 TYPE: DNA
 ORGANISM: Artificial Sequence

Query Match	7.0%;	Score 32;	DB 6;	Length 1355;
Best Local Similarity	25.4%;	Pred. No. 0.49;		
Matches	59;	Conservative	64;	Mismatched 100

[illegible]

RESULT 7
US-10-767-471-10809/c

```

? sequence 10809, Application US/10767471
? GENERAL INFORMATION:
? APPLICANT: CARGILL, Michele et al
? TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
? FILE OF INVENTION: RHUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
? CURRENT APPLICATION NUMBER: US/10/767,471
? CURRENT FILING DATE: 2004-01-30
? NUMBER OF SEQ ID NOS: 50231
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 10809
? LENGTH: 290892
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (1)...(290892)
? OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

```

Query Match	6.8%;	Score 30.8;	DB 6;	Length 290892;
Best Local Similarity	46.3%;	Pred. No. 16;		
Matches 101; Conservative	0;	Mismatches 117		

[illegible]

RESULT 8

Query Match Score 30; DB 6; Length 372;
Best Local Similarity 43.1%; Pred. No. 1.3;
Matches 144; Conservative 0; Mismatches 190; Indels 0

Oy 62 GTGTTGTTCCTCAAGTACGGCCGCCGGTAACAACGGTGCGCGGTATATATACGGCC 121
Db 340 GTATATGATGAATGAATGCGTGATATAGTATGATGATGATGAATGATGCGTATATATGATGATG 281

QY 122 CAATTCTGAGCTGAACATTACCACTGCGGTAATCTCTGACTTCTTGCAGAA 181
DB 280 ATGATGCTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 221
QY 182 CTGATCCCGTAACCTGACTGATGATGATGATGATGATGATGATGATGATGATG 241
DB 220 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 161
QY 242 TTGCTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
DB 160 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 101
QY 302 CTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
DB 100 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 41
QY 362 GCAACGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 395
DB 40 ATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7

RESULT 12

PCT-US04-05654-1019
Sequence 1019, Application PC/TUS0405654
GENERAL INFORMATION:

APPLICANT: Sherman, Bradley K
APPLICANT: Riechmann, Jose Luis
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Heide, Jacqueline E
APPLICANT: Haake, Volker
APPLICANT: Creelman, Robert A
APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc J
APPLICANT: Reuber, T. Lynne
APPLICANT: Keddie, James
APPLICANT: Dubeil III, Arnold N
APPLICANT: Pineda, Omaira
APPLICANT: Repetti, Peter
APPLICANT: Century, Karen
APPLICANT: Guterson, Neal
APPLICANT: Yu, Guo-Liang
APPLICANT: Brown, Pierre E
APPLICANT: Kumamoto, Roderick W
APPLICANT: Pilgrim, Marsha L
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
FILE REFERENCE: MB1-0047 PCT
CURRENT APPLICATION NUMBER: PCT/US04/05654
PRIOR FILING DATE: 2004-03-03
PRIOR APPLICATION NUMBER: 10/374,780
PRIOR FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: 10/675,852
NUMBER OF SEQ ID NOS: 2950
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1019
LENGTH: 992
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
FEATURE:
NAME/KEY: misc feature
LOCATION: (978)..(984)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (987)..(987)
OTHER INFORMATION: n is a, c, g, or t
PCT-US04-05654-1019

Query Match 6.6%; Score 30; DB 1; Length 992;
Best Local Similarity 61.5%; Pred. No. 2.1;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 42 CTCGGTAGGCGCTGCGAGGTGTTCTCTCACTACCGGGGGCGGTAACCAAGGTGG 101
DB 324 CTATGCGGGGGGACGAGAGGTGTTGAGGTGATGATGATGATGATGATGATG 383
QY 102 TGGCGGTAAATATAGCGG 119
DB 384 TGGCGGTGCTATGTTGG 401

RESULT 13

US-10-021-698A-796/c
Sequence 796, Application US/10021698A
GENERAL INFORMATION:

APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN EERDEWEGH, PAUL
APPLICANT: DUPUIS, JOSEF
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
APPLICANT: PANDIT, SUNIL
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044US1
CURRENT APPLICATION NUMBER: US/10/021,698A
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
PRIOR FILING DATE: 2000-06-14
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: PatentIn 2.1
SEQ ID NO 796
LENGTH: 9278
TYPE: DNA
ORGANISM: Homo sapiens
US-10-021-698A-796

Query Match 6.6%; Score 30; DB 6; Length 9278;
Best Local Similarity 46.3%; Pred. No. 5.9;
Matches 99; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 46 GGTAAGCGCTGCGAGGTGTTCTCTCACTACCGGGGGCGGTAACCAAGGTGG 105
DB 359 GGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 106 GGTAATATAGCGGCGCAATTTCTGAGCTGAACATTTACAGTACGTCGCGTAACCT 165
DB 299 GATGCTGACGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
QY 166 GCACTTCTCTGCAACTGATGATGATGATGATGATGATGATGATGATGATGATG 225
DB 239 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 226 GGTAATGCTGAGATGTTGCTGAGGCTCAGATG 259
DB 179 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 146

RESULT 14

PCT-US04-02242-43
Sequence 43, Application PC/TUS0402242
GENERAL INFORMATION:

APPLICANT: Diversa Corporation
APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING AND
FILE REFERENCE: 56446-20118_40
CURRENT APPLICATION NUMBER: PCT/US04/02242
PRIOR FILING DATE: 2004-01-26
PRIOR APPLICATION NUMBER: 60/442,794
PRIOR FILING DATE: 2003-01-24
NUMBER OF SEQ ID NOS: 224
SOFTWARE: PatentIn version 3.1

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456
Sequence: 1 atgaaaaaacattgtctatc.....ctgcgtgacacacgttaa 456

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patente NA Main:

```
1: /cgn2_6/prodata/2/pna/US06_PCTUS_COMB.seq:*
2: /cgn2_6/prodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/prodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/prodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/prodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/prodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/prodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/prodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/prodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/prodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/prodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/prodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/prodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/prodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/prodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/prodata/2/pna/US092_COMB.seq:*
17: /cgn2_6/prodata/2/pna/US093_COMB.seq:*
18: /cgn2_6/prodata/2/pna/US093B_COMB.seq:*
19: /cgn2_6/prodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/prodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/prodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/prodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/prodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/prodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/prodata/2/pna/US095A_COMB.seq:*
26: /cgn2_6/prodata/2/pna/US095B_COMB.seq:*
27: /cgn2_6/prodata/2/pna/US095C_COMB.seq:*
28: /cgn2_6/prodata/2/pna/US095D_COMB.seq:*
29: /cgn2_6/prodata/2/pna/US096_COMB.seq:*
30: /cgn2_6/prodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/prodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/prodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/prodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/prodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/prodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/prodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/prodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/prodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/prodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/prodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/prodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/prodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/prodata/2/pna/US100A_COMB.seq:*
```

```
44: /cgn2_6/prodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/prodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/prodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/prodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/prodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/prodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/prodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/prodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/prodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/prodata/2/pna/US105_COMB.seq:*
54: /cgn2_6/prodata/2/pna/US106_COMB.seq:*
55: /cgn2_6/prodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/prodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/prodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/prodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/prodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/prodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/prodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/prodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/prodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/prodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/prodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/prodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/prodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/prodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/prodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/prodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/prodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/prodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/prodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/prodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/prodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/prodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/prodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/prodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/prodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/prodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/prodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/prodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/prodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/prodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/prodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/prodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/prodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/prodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/prodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/prodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/prodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/prodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/prodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/prodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/prodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/prodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/prodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/prodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/prodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/prodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/prodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/prodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/prodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/prodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/prodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/prodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/prodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/prodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/prodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/prodata/2/pna/US6053_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Description
Sequence 4, Appl
Sequence 2, Appl
Sequence 171, Appl
Sequence 171, Appl
Sequence 171, Appl
Sequence 80, Appl
Sequence 82, Appl
Sequence 14, Appl
Sequence 15700, A
Sequence 15700, A
Sequence 3221, Ap
Sequence 5263, Ap
Sequence 13, Appl
Sequence 17017, A
Sequence 82785, A
Sequence 3919, Ap
Sequence 3919, Ap
Sequence 25, Appl
Sequence 51735, A
Sequence 85430, A
Sequence 273, Appl
Sequence 156, App
Sequence 156, App
Sequence 156, App
Sequence 156, App
Sequence 87989, A
Sequence 5135, Ap
Sequence 72, Appl
Sequence 1052, Ap
Sequence 72, Appl
Sequence 3079, Ap
Sequence 28400, A
Sequence 10097, A
Sequence 80300, A
Sequence 47103, A
Sequence 22694, A
Sequence 64656, A
Sequence 40033, A
Sequence 20466, A
Sequence 66935, A
Sequence 15616, A
Sequence 19909, A
Sequence 5821, Ap
Sequence 76, Appl

ALIGNMENTS

```

RESULT 1
US-09-543-407-4
Sequence 4, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FILTRAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS
FILE REFERENCE: 920043,406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
SOFTWARE: FaalSED for Windows Version 4.0
SEQ ID NO 4
LENGTH: 456
TYPE: DNA
ORGANISM: E. COLI

```

US-09-543-407-4

Query Match	100.0%;	Score 456;	DB 23;	Length 456;
Best Local Similarity	100.0%;	Pred. No. 9e-125;		
Matches 456;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

2Y 1 ATGAATAACAATTGTTATTATGATGTTAACATACTGGGTCCGCTGGGATTGCAGCC 60

Db 1 ATGAATAACAAATTGTTATTATGATGTTAACATACTGGGTGCCCTGGGATTGCAGCC 60

61 GCAGCAGGTTATGATTTAGCTAATTCAGAATATACTTCGGGTAATGAATTGAGTAAG 120

61 GCAGCAGGTATGATTAGCTAATTCAGAATATACTTCGGGTAATGCAATTCAGTAAG 120

121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGGACTAATAATAGTCTCAGTTA 180

121 TCTTCATTTAATCAGGCAGCCATAATTGGTCAAGCTGGACTAATAATAGTGTCTCAGTTA 180

181 CGCAGGAGGCTCAAACTTTGGCGTTGTTCGCAGAAGTAGTAGCAACCGGCA 240

181 CGCAGGAGGC TC AAAACTTTGGCGGTGTTGCCAAGAAGTAGTAGCAACCGGCA 240

241 AAGATTCACCAAGAGATTAACTTCATATATTGATCAGCGGGCAGTGCAC 300

241 AAGATGACCAGACGAGATTAACCTTGATATATTGATCAGGGGGCAGTGCAC 300

301 GATGCCAGTATTCGCAAGTGGCTATGGTATACTCGATGATTATCCAGAAAGTTCT 360

351 GAGGCCAGATTTCCGAAGTCTATGGTAACTCCGATGATTCCAGAAGTTCT 360

CCATGCTTATTTCACAGTAAGGACTCAAAAACGGCAATTGTGAGTCAGAGA 420

CCCTCCTTTCCTTTATACACGTAAGTACTCAAAAACGGCAATTGTAGTGCAGAGA 420

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

[illegible]

RESULT 2

Sequence 2, Application US/09543407

APPLICANT: white, Aaron P.

APPLICANT: Collinson, S. Karen

1. TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

FILE REFERENCE: 920043.406

CURRENT FILING DATE: 2000-04-05

```

; SOFTWARE: FastSEQ for windows Version 4.0
; SEQ ID NO 3

```

```

; LENGTH: 456
TYPE: 0000

```

ORGANISM: *Salmonella enteritidis*

Answer Match

Best Local Similarity 83.1%; Pred. No. 4.7e-88;
Matches 379; Concordance 0.91; Winsteps 27

[illegible]

b

ATGAAAAACAAATTCTTATTTATTCATCATAAGTAATCAATCATCCATCGAATGGGAATG

61 GCGAGGCTTATGGATTTCACATAAATCAACAAGAATTAATCATTTAACTCAGCCAGCTTG

[illegible]

121 TCCTCAATTAAATCAGGCGACCATTAATTGTGGTGAAGCTTGCCACTTCATTTCTGCCT

```

Db      121 TCTTCATTATACGCGGCATTATGTCAAGTCGACGAGTAATAGTCCAGACTA 180
Qy      181 CGGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACACCGGGCA 240
Db      181 CGCGAGGAGGATCAAACTATGTGCTTATTTCAAGAAGAGGAAATATCGGGCG 240
Qy      241 AAGATTGACGACGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAC 300
Db      241 AAGTGCACGAGGAGGAGATTAATTAATCTTGATATATGAGCAACCGGCAATGCCAC 300
Qy      301 GATCCAGATTTTGCAGAGGTGCTTATGTAATATCTGGATGATTAATCAGAAAGTTCT 360
Db      301 GATCCAGATTTTGCAGAAAGGCTTAAGGTATAGTGCAGCTATTAATCAGAAAGTTCT 360
Qy      361 GGTATTAAGCAATATTTACACAGTATGTAATCTCAAAAACGCGCAATTGATGACAGA 420
Db      361 GGAATTAAGGCAATATTTACACAGTATGTAATCTCAAAAACGAGATTGTAGTACAGAAA 420
Qy      421 CAGTCGCAATGCTATTGCGGTGACACAGCTTAA 456
Db      421 CAGTCGCAATGCTATTGCGGTGACACAGCTTAA 456

```

RESULT 3

```

US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252.691B
; PRIORITY FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 171
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-171

```

```

Query Match      63.6%; Score 290.2; DB 17; Length 561;
Best Local Similarity 77.4%; Pred. No. 2.5e-75;
Matches 352; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

```

Qy      1 ATGAAAAACAATTTGTTATTTATGTTAAACAATCTGGGTGCGCTGGAGTTGCAGCC 60
Db      106 ATGAAAAACAACGTTGTTATGATGTTTACATTACTGGGTGCGCTGGTTGTATC 165
Qy      61 GCACAGGTTATGATTTAGCTAATTCAGATTAATTAATCTGGGTGAGTAATGATGATAG 120
Db      166 GCAGGTATTCAGATTAGCCAGTTCTGAATATTAATTTGCGATTAATGAATTAAGTAA 225
Qy      121 TCTTCATTATACGAGGACCATATTTGTCAGCTGAGCTGGAATTAATAGTCTCAGTTA 180
Db      226 GCTTCATACATACGAGGACCATATTTGTCAGAGGTTTCAAGAAATTAATTCAGATGA 285
Qy      181 CGGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACACCGGGCA 240
Db      286 CGCAGAGAGGCTTCAAACTGTCGTTATTTCTCAGAGGCGGGGAATTAACCGGCG 345
Qy      241 AAGATTGACGACGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 300
Db      346 AATGTTATCAGTCAGAGGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 405
Qy      301 GATCCAGATTTTGCAGAGGTGCTTATGTAATATCTGGATGATTAATCAGAAAGTTCT 360
Db      406 GATCCAGATTTTGAAGGCGGCTTTGGCAACCGCGCATGATTAATCAGAAAGGCTCG 465
Qy      361 GGTATTAAGCAATATTTACACAGTATGTAATCTCAAAAACGCGCAATTGATGACAGA 420
Db      466 GGTAAAGGCGGAGATTAATTAACAGTATGTAATCTCAAAAACGAGATTGTAGTACAGAGA 525

```

```

Qy      421 CAGTCGCAATGCTATTGCGGTGACACAGCTTA 455
Db      526 CAGTCGCAATGCTATTGCGGTATTCAAGCTTA 560

```

RESULT 4

```

US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252.691C
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094.145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074.787
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-171

```

```

Query Match      63.6%; Score 290.2; DB 17; Length 561;
Best Local Similarity 77.4%; Pred. No. 2.5e-75;
Matches 352; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

```

Qy      1 ATGAAAAACAATTTGTTATTTATGTTAAACAATCTGGGTGCGCTGGAGTTGCAGCC 60
Db      106 ATGAAAAACAACGTTGTTATGATGTTTACATTACTGGGTGCGCTGGTTGTATC 165
Qy      61 GCACAGGTTATGATTTAGCTAATTCAGATTAATTAATCTGGGTGAGTAATGATGATAG 120
Db      166 GCAGGTATTCAGATTAGCCAGTTCTGAATATTAATTTGCGATTAATGAATTAAGTAA 225
Qy      121 TCTTCATTATACGAGGACCATATTTGTCAGCTGAGCTGGAATTAATAGTCTCAGTTA 180
Db      226 GCTTCATACATACGAGGACCATATTTGTCAGAGGTTTCAAGAAATTAATTCAGATGA 285
Qy      181 CGGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACACCGGGCA 240
Db      286 CGCAGAGAGGCTTCAAACTGTCGTTATTTCTCAGAGGCGGGGAATTAACCGGCG 345
Qy      241 AAGATTGACGACGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 300
Db      346 AATGTTATCAGTCAGAGGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 405
Qy      301 GATCCAGATTTTGCAGAGGTGCTTATGTAATATCTGGATGATTAATCAGAAAGTTCT 360
Db      406 GATCCAGATTTTGAAGGCGGCTTTGGCAACCGCGCATGATTAATCAGAAAGGCTCG 465
Qy      361 GGTATTAAGCAATATTTACACAGTATGTAATCTCAAAAACGCGCAATTGATGACAGA 420
Db      466 GGTAAAGGCGGAGATTAATTAACAGTATGTAATCTCAAAAACGAGATTGTAGTACAGAGA 525
Qy      421 CAGTCGCAATGCTATTGCGGTGACACAGCTTAA 455
Db      526 CAGTCGCAATGCTATTGCGGTATTCAAGCTTA 560

```

RESULT 5

```

US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10/417.886
; PRIORITY FILING DATE: 2003-04-17

```

PRIOR APPLICATION NUMBER: US/09/252,691C
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,145
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/074,787
PRIOR FILING DATE: 1998-02-18
NUMBER OF SEQ ID NOS: 11326
SEQ ID NO 171
LENGTH: 561
TYPE: DNA
ORGANISM: Enterobacter cloacae
US-10-417-886-171

Query Match 63.6%; Score 290.2; DB 51; Length 561;
Best Local Similarity 77.4%; Pred. No. 2.5e-75;
Matches 352; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1 ATGAAAAAATAATTTATTTATGATGTTTAACTAGTGGTGGCTGGAGTGCAGCC 60
DB 106 ATGAAAAAATAATTTATTTATGATGTTTAACTAGTGGTGGCTGGAGTGCAGCC 165
QY 61 GCAGCAGTATGATTTAGCTATTTAGAAATTAATTTGCGGTAATGAATTGAGTAG 120
DB 166 GCAGCAGTATGATTTAGCTATTTAGAAATTAATTTGCGGTAATGAATTGAGTAG 225
QY 121 TCTTCATTTATGACGACGCAATTTGTCAGAGCGGGAATAATGATGCTAGTAA 180
DB 226 GCTTCATACATGACGACGCAATTTGTCAGAGCGGGAATAATGATGCTAGTAA 285
QY 181 CGCAGCAGGCGCTCAAACTTTGGCGGTTGGCGAAGAGTAGTAGCAACCGGCA 240
DB 286 CGCAGCAGGCGCTCAAACTTTGGCGGTTGGCGAAGAGTAGTAGCAACCGGCA 345
QY 241 AAGATTGACACAGAGAGATATTAACCTGATATTTATGACGCGGCGAGTGCAC 300
DB 346 AATGTTGATCAGTACGACGCAATTTGTCAGAGCGGGAATAATGATGCTAGTAA 405
QY 301 GATGCGATTTTGCAGAGTCTTATGATATCTGCGATGATATTCAGAAAGTTCT 360
DB 406 GATGCGATTTTGCAGAGTCTTATGATATCTGCGATGATATTCAGAAAGTTCT 465
QY 361 GGTATTAAGCAAAATATTAACAGATAGTACTCAAAAAAGCAATTGATGACAGAGA 420
DB 466 GGTATTAAGCAAAATATTAACAGATAGTACTCAAAAAAGCAATTGATGACAGAGA 525
QY 421 CAGTGCAGATGCTATTTGCGGTGACACAGCTTA 455
DB 526 CAGTGCAGATGCTATTTGCGGTGACAGCTTA 560

RESULT 6
US-10-146-492B-80
Sequence 80, Application US/10146492B
GENERAL INFORMATION:
APPLICANT: MMG Biotech AG
TITLE OF INVENTION: Biochip
FILE REFERENCE: REN-14043
CURRENT APPLICATION NUMBER: US/10/146,492B
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: EP 0111279.5
NUMBER OF SEQ ID NOS: 12816
SEQ ID NO 80
LENGTH: 100
TYPE: DNA
ORGANISM: Escherichia coli K-12 MG1655
FEATURE:
OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
PUBLICATION INFORMATION:
RELEVANT RESIDUES: 39-58
US-10-146-492B-80

Query Match 21.9%; Score 100; DB 45; Length 100;

Best Local Similarity 100.0%; Pred. No. 6.7e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 266 ACCTTGATATATGATCAGCGGCGGAGTGCACAGATGATTTGGCAAGTCTT 325
DB 1 ACCTTGATATATGATCAGCGGCGGAGTGCACAGATGATTTGGCAAGTCTT 60
QY 326 ATGGTATATCTGCATATATTCAGAAAGTTCTGTTAA 365
DB 61 ATGGTATATCTGCATATATTCAGAAAGTTCTGTTAA 100

RESULT 7
US-10-146-492B-81
Sequence 81, Application US/10146492B
GENERAL INFORMATION:
APPLICANT: MMG Biotech AG
TITLE OF INVENTION: Biochip
FILE REFERENCE: REN-14043
CURRENT APPLICATION NUMBER: US/10/146,492B
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: EP 0111279.5
NUMBER OF SEQ ID NOS: 12816
SEQ ID NO 81
LENGTH: 100
TYPE: DNA
ORGANISM: Escherichia coli K-12 MG1655
FEATURE:
OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
PUBLICATION INFORMATION:
RELEVANT RESIDUES: 38-57
US-10-146-492B-81

Query Match 21.9%; Score 100; DB 45; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.7e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GTTGTGGCGAAGAGTAGTAGCAACCGGCAAGATGACAGAGGATATTAAC 267
DB 1 GTTGTGGCGAAGAGTAGTAGCAACCGGCAAGATGACAGAGGATATTAAC 60
QY 268 CTTGATATATGATCAGCGGCGGAGTGCACAGATGCCA 307
DB 61 CTTGATATATGATCAGCGGCGGAGTGCACAGATGCCA 100

RESULT 8
US-10-146-492B-82
Sequence 82, Application US/10146492B
GENERAL INFORMATION:
APPLICANT: MMG Biotech AG
TITLE OF INVENTION: Biochip
FILE REFERENCE: REN-14043
CURRENT APPLICATION NUMBER: US/10/146,492B
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: EP 0111279.5
NUMBER OF SEQ ID NOS: 12816
SEQ ID NO 82
LENGTH: 100
TYPE: DNA
ORGANISM: Escherichia coli K-12 MG1655
FEATURE:
OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
PUBLICATION INFORMATION:
RELEVANT RESIDUES: 57-76
US-10-146-492B-82

Query Match 21.9%; Score 100; DB 45; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.7e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

APPLICANT: Kofler, Janette
APPLICANT: Labat, Ivan
APPLICANT: Lee, Won-Jae
APPLICANT: Lomelli, Michelle
APPLICANT: Nguyen, Hong
APPLICANT: Nguyen, Linh
APPLICANT: Nguyen, Lynne
APPLICANT: Nguyen, Phuong
APPLICANT: Ngira, Margie
APPLICANT: Ojeda, Jesse
APPLICANT: Palencia, Servando
APPLICANT: Ralei, Fariba
APPLICANT: Randhwa, Gurpreet
APPLICANT: Sabourieh, Hannah
APPLICANT: Sidhu, Navjivan
APPLICANT: Smith, Benjamin
APPLICANT: Smythe, Ashleigh
APPLICANT: Tkach, Joe
APPLICANT: Tulpule, Mukul
APPLICANT: Verna, Ron
APPLICANT: Wachter, Adam
APPLICANT: Wu, James
APPLICANT: Yim, Kenneth
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 754C1P
CURRENT APPLICATION NUMBER: US/09/515,694
CURRENT FILING DATE: 2000-02-29
EARLIER APPLICATION NUMBER: 09/332,782
EARLIER FILING DATE: 1999-06-14
EARLIER APPLICATION NUMBER: 09/181,430
EARLIER FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 21027
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 15700
LENGTH: 391
TYPE: DNA
ORGANISM: Homo sapiens
US-09-515-694-15700

Query Match
Best Local Similarity 8.4%; Score 38.2; DB 21; Length 391;
Pred. No. 2.7;
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 309 TATTTCCAGAGGCTTATGTATCTGCGATGATATTCAGAAAGTCTGTGATTA 368
DB 277 TTTCGGCGCAAGCTGGTGTCTGCTGACCTGATATTCGCAATTAATCTGGAATGT 336
QY 369 AGCAATATATACAGATGATGCTCAAAAACGGCAATGATGAGAGAGAG 423
DB 337 AGCCAGACTGCAATATATATGACATATATCAGACACTTCTGTAGCATATCTG 391

RESULT 12
US-10-431-652-3221
Sequence 3221, Application US/10431652
GENERAL INFORMATION:
APPLICANT: Breton, Gary L.
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: PATH03-08
CURRENT APPLICATION NUMBER: US/10/431,652
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: US 09/328,352
PRIOR FILING DATE: 1999-06-04
PRIOR APPLICATION NUMBER: US 60/088,701
PRIOR FILING DATE: 1998-06-09
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 3221
LENGTH: 4860
TYPE: DNA
ORGANISM: Acinetobacter baumannii

US-10-431-652-3221

Query Match
Best Local Similarity 8.4%; Score 38.2; DB 51; Length 4860;
Pred. No. 6.3;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 106 AATGAATGAGTAGCTTCAATTTATTCAGGAGCCATTAATGTCAGCTGGACTAAT 165
DB 3445 AATACTATTGGTAAATTAATTAAGTACTGATGAGCAGATTAATCTTAAGAGATGCTGAG 3504
QY 166 AATAGTCTCAGTTACGAGGAGGCTCAAACTTTGGCGGTGTTGCGCAAGAGT 225
DB 3505 ATTTCAACTATTATGTTTGGAGGCANAGATGATTCAGATGCTGTACGATGAT 3564
QY 226 AGTAGCAACCGGCAAGATTGACGACAGAGATTAATACCTGATATATGATCAG 285
DB 3565 TATCTGTATGATGATGATGAGACGATTAATGTTCTTAATACAGGCTCTGATTTTG 3624
QY 286 GCGGAGTGCACAGATGCCAGTATTTGCAAGGCTTATGTAATCTGGATGAT 345
DB 3625 TATGTTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3684
QY 346 ATCCAGAAAGTCTGTGATTAATTAAGCAATATATACAGATG 388
DB 3685 TTACAGATGCAAGCTGGATGATGATGATGATGATGATGATGATGATGATGATGAT 3727

RESULT 13
US-09-887-272A-5263/C
Sequence 5263, Application US/09887272A
GENERAL INFORMATION:
APPLICANT: Hou, Yu-Ming
APPLICANT: Quan, Sheng
APPLICANT: Chang, Hui-Song
APPLICANT: Zhu, Tong
APPLICANT: Whitam, Steve
APPLICANT: Goff, Steve
APPLICANT: Glazebrook, Jane
APPLICANT: Chen, Wenguiang
APPLICANT: Katagiri, Fumaki
APPLICANT: Xie, Zhiyi
APPLICANT: Tao, Yi
APPLICANT: Zou, Guangzhou
APPLICANT: Cooper, Bret
TITLE OF INVENTION: PLANT GENES INVOLVED IN DEFENSE AGAINST
FILE REFERENCE: PATH03-08
CURRENT APPLICATION NUMBER: US/09/887,272A
CURRENT FILING DATE: 2001-06-23
PRIOR APPLICATION NUMBER: 60/213,634
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/214,926
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/261,320
PRIOR FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 60/264,353
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/273,879
PRIOR FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: 09/887,271
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 6813
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5263
LENGTH: 2000
TYPE: DNA
ORGANISM: Oryza sativa
US-09-887-272A-5263

Query Match
Best Local Similarity 8.3%; Score 37.8; DB 36; Length 2000;
Pred. No. 6.2;
Matches 36; Conservative 121; Mismatches 118; Indels 0; Gaps 0;

```
Qy 21 TATGATGTAACAATACTGGGCGCTGGGATTCAGCCGAGGAGTTAGATTGAC 80
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 950 WAAGRRMRMAWCTCCMKMKMKMTSCMMWKWRWSCTYMMCAMRYAAYAMRRRMT 891
Qy 81 TAAATGAGATATAAATTCGGCGTAATGAATGAGTCTTCAATTAATCAGGACG 140
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 890 YKMSRMRMYMTWTKAMTMTCMCAKMYMATGATMMMYTYTTCYATTCCKCKYKM 831
Qy 141 CATATGTCAGCTGGAGCTAATATAGTCTCAGTTACGGCAGGAGGCTCAAACT 200
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 830 AMTYMTWTMACAMRATSWMRAMGMRKRYKMRAYMRWRCKAGMARMRKRYRMK 771
Qy 201 TTGGCGGTTGTCGCAAGAAGTAGACACCGGCAAGAATTGACAGACAGAGA 260
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 770 WKATATRYWGMAMTMMWSMRKMSYMMSGMRMRMSARVSRMKCATKXIASAKM 711
Qy 261 TTATTAACCTTCGATATATGATTCAGCGCGGAGTG 295
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 710 TKRARSYRVRMRWYMRKKGWTVRYRVRSCRMTR 676
```

RESULT 14
US-09-543-407-13

```
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; CURRENT APPLICATION NUMBER: US/09/543,407
; FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13
```

Query Match 8.2%; Score 37.4; DB 23; Length 456;
Best Local Similarity 51.5%; Pred. No. 5;
Matches 86; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

```
Qy 277 ATTGATCAGGCGGCGAGTCCAGATGCGGATTTTCGCAAGTCTTATGTAATCT 336
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 208 ATTGATCAGGCGGCGGTTATGTTAGCGCGCCGATGAGCGCGAGGTGCGGATATAGTACT 267
Qy 337 GCGATGATTATTCAGAAAGTTCGTGTATTAAGCAATATTAACAGATGATGCTCAA 396
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 268 ATTGAATGATCTCGAATGTTTTCAGAAATTAATGACCATGACAGTGAACGCTTAA 327
Qy 397 AAAACGGCAATTTAGTGCAGACAGTCCGCAATGCGTATTCGCT 443
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 328 AACTCCGATATTACTGTGCGCAATATGATGATGCTGTTACCGGT 374
```

RESULT 15
US-09-949-016-17017/c
; Sequence 17017, Application US/09949016
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CU001307
; CURRENT APPLICATION NUMBER: US/09/949, 016
; CURRENT FILING DATE: 2000-04-14

```
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17017
; LENGTH: 68667
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(68667)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17017
```

Query Match 8.2%; Score 37.4; DB 39; Length 68667;
Best Local Similarity 58.6%; Pred. No. 27;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy 327 TGGTAATGCTGCGATGATTAATCCAGAAAGTTCGTGTATTAAGCAATATTAACAGTA 386
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 17497 TGCTCATTCGTGTATTAATTTCTTAAGTGTAGGAAATTAATTAATTAATTA 17438
Qy 387 TGGTACTCAAAAAGCGCAATTTGATGTCAGACAGTCCGCAATGCGCTAT 437
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 17437 TGATATGAGAAATGACACATGTGTGTATGTTACTTCCAAATTAATCTT 17387
```

Search completed: March 18, 2004, 02:46:53
Job time : 2927.24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456

Sequence: 1 atgaaatacaatgttctatc.....ctgcgtgacacacgttaa 456

Scoring table: IDENTITY_NUC

Searched: Gap0 10.0, Gapext 1.0

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending Patents NA New:
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.4	8.2	65558	US-10-765-790-103	Sequence 103, App
2	35	7.7	675	US-09-741-873C-3	Sequence 73, Appl
3	34.4	7.5	519599	US-10-765-790-73	Sequence 73, Appl
4	34.2	7.5	887	US-10-767-701-9911	Sequence 9911, App
5	33.8	7.4	1149	US-10-771-241-139	Sequence 139, App
6	33.4	7.3	38719	US-10-765-790-43	Sequence 43, Appl
7	32.2	7.1	261922	US-10-767-471-10653	Sequence 10653, A
8	31.6	6.9	512	US-10-767-701-11198	Sequence 11198, A
9	31.4	6.9	525	US-10-417-884A-2325	Sequence 2325, Ap
10	31.4	6.9	1245	US-60-546-745-36	Sequence 36, Appl
11	31.4	6.9	53332	US-10-766-065-3	Sequence 3, Appl1
12	30.8	6.8	201	US-10-767-471-8836	Sequence 8836, Ap
13	30.8	6.8	201	US-10-767-471-34493	Sequence 34493, A
14	30.8	6.8	1227	US-10-767-471-501	Sequence 501, App
15	30.8	6.8	76573	US-10-767-471-10718	Sequence 10718, A
16	30.4	6.7	1227	US-10-417-884A-2544	Sequence 2544, Ap
17	30.2	6.6	3529	US-10-779-543-9918	Sequence 9918, Ap
18	30.2	6.6	1790242	US-10-767-471-10805	Sequence 10805, A
19	29.8	6.5	600	US-60-545-213-3515	Sequence 3515, Ap
20	29.8	6.5	600	US-60-545-213-7787	Sequence 7787, Ap
21	29.6	6.5	201	US-10-767-471-8837	Sequence 8837, Ap
22	29.6	6.5	201	US-10-767-471-34439	Sequence 34439, A
23	29.6	6.5	40000	US-10-767-471-10911	Sequence 10911, A
24	29.6	6.5	122673	US-10-765-790-33	Sequence 33, Appl
25	29.4	6.4	2806	US-10-021-698A-2111	Sequence 2111, Ap
26	29.4	6.4	74677	US-10-767-471-10891	Sequence 10891, A

27	29.2	6.4	1080	US-10-451-467A-619	Sequence 619, App
28	29.2	6.4	134134	US-10-767-471-10743	Sequence 10743, A
29	29.2	6.4	189268	US-10-021-698A-697	Sequence 697, App
30	29.2	6.4	290892	US-10-767-471-10809	Sequence 10809, A
31	29	6.4	4447	US-10-767-701-18004	Sequence 18004, A
32	29	6.4	532	US-10-767-701-17618	Sequence 17618, A
33	29	6.4	13948	US-10-767-471-10858	Sequence 10858, A
34	29	6.4	16201	US-10-767-471-10705	Sequence 10705, A
35	29	6.4	45268	US-60-548-091-5603	Sequence 5603, Ap
36	28.8	6.3	1829	US-10-775-169-326	Sequence 326, App
37	28.8	6.3	1914	US-10-767-701-13618	Sequence 13618, A
38	28.8	6.3	20767	US-10-100-663-12012	Sequence 12012, A
39	28.6	6.3	408	US-10-417-884A-793	Sequence 793, App
40	28.6	6.3	1056	US-10-781-014-285	Sequence 285, App
41	28.6	6.3	1036	US-10-781-014-449	Sequence 449, App
42	28.6	6.3	78785	US-09-978-167A-3	Sequence 3, Appl1
43	28.6	6.3	100374	US-60-548-091-5675	Sequence 5675, Ap
44	28.6	6.3	189817	US-10-767-471-10675	Sequence 10675, A
45	28.4	6.2	477	US-10-779-543-18788	Sequence 18788, A

ALIGNMENTS

```
RESULT 1
US-10-765-790-103/C
Sequence 103, Application US/10765790
GENERAL INFORMATION:
APPLICANT: Bayer Healthcare LLC
APPLICANT: Beard, Chris
APPLICANT: Burgess, Chris
APPLICANT: Gannon, Allison
APPLICANT: Harvey, Jeanne
APPLICANT: Lechner, John F.
APPLICANT: Li, Zheng
TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
FILE REFERENCE: 1657/2035
CURRENT APPLICATION NUMBER: US/10/765,790
CURRENT FILING DATE: 2004-01-27
PRIOR APPLICATION NUMBER: US 10/737,082
PRIOR FILING DATE: 2003-12-16
NUMBER OF SEQ ID NOS: 300
SOFTWARE: PatentIn version 3.2
SEQ ID NO 103
LENGTH: 65558
TYPE: DNA
ORGANISM: Homo sapiens
US-10-765-790-103

Query Match      8.2%; Score 37.4; DB 6; Length 65558;
Best Local Similarity 58.6%; Pred. No. 0.2;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Cy 327 TGGTAATACGCGATGATTAATCCGAAGGTTGTGTAATTAACCAATATTACACAGTA 386
Db 16503 TGCATCTTCGTGTGTAATATTCCTTAAGTGTAAGGAAATTAATTAATTAATTA 16444
      387 TGGTACTCAAAAACGCGCAATTGATGTCAGAGACAGTCGCAATGCGCTAT 437
Db 16443 TGATATGAGAAATAGCAGACAGTGTGCTGATGTACTTCCAAATTACTTT 16393

RESULT 2
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012869-084
CURRENT APPLICATION NUMBER: US/09/741,873C
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
```

;; PRIOR FILING DATE: 1998-05-06
;; PRIOR APPLICATION NUMBER: US 08/978,878
;; PRIOR FILING DATE: 1997-11-26
;; PRIOR APPLICATION NUMBER: US 07/347,189
;; PRIOR FILING DATE: 1989-05-04
;; PRIOR APPLICATION NUMBER: US 07/789,437
;; PRIOR FILING DATE: 1991-11-06
;; PRIOR APPLICATION NUMBER: US 07/970,846
;; PRIOR FILING DATE: 1992-11-03
;; PRIOR APPLICATION NUMBER: US 08/187,865
;; PRIOR FILING DATE: 1994-01-28
;; PRIOR APPLICATION NUMBER: US 08/318,519
;; PRIOR FILING DATE: 1994-10-05
;; NUMBER OF SEQ ID NOS: 11
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 3
;; LENGTH: 675
;; TYPE: DNA
;; ORGANISM: Escherichia coli
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (83)..(538)
US-09-741-873C-3

Query Match
Best Local Similarity 7.7%; Score 35; DB 5; Length 675;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 416 AGAGACAGTCGCAAAATGGCTATTCGGCTGACACAA 450
Db 1 AGAGACAGTCGCAAAATGGCTATTCGGCTGACACAA 35

RESULT 3
US-10-765-790-73/c
;; Sequence 73, Application US/10765790
;; GENERAL INFORMATION:
;; APPLICANT: Bayer Healthcare LLC
;; APPLICANT: Beard, Chris
;; APPLICANT: Burgess, Chris
;; APPLICANT: Gannon, Allison
;; APPLICANT: Harvey, Jeanne
;; APPLICANT: Lechner, John F.
;; APPLICANT: Li, Zheng
;; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;; FILE REFERENCE: 1657/2035
;; CURRENT APPLICATION NUMBER: US/10/765,790
;; CURRENT FILING DATE: 2004-01-27
;; PRIOR APPLICATION NUMBER: US 10/737,082
;; PRIOR FILING DATE: 2003-12-16
;; NUMBER OF SEQ ID NOS: 300
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 73
;; LENGTH: 519599
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-765-790-73

Query Match
Best Local Similarity 7.5%; Score 34.4; DB 6; Length 519599;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
Qy 263 ATAACTTGATATATGATCAGCGGCGGACATGCCACTATTTCGCAAGTG 322
Db 332273 ATAAATTTGATTTTGTGATCAGTACGATGGGCTATCTCTGACAGCATTTGCAAAAGAG 332214
Qy 323 CTATAGTAATAGTGCATGATTAATTCAGAAA 354
Db 332213 CTATAGTATACAGAAAATATACCAAAA 332182

RESULT 4
US-10-767-701-9911

;; Sequence 9911, Application US/10767701
;; GENERAL INFORMATION:
;; APPLICANT: Kovalic, David K.
;; APPLICANT: Zhou, Yihua
;; APPLICANT: Cao, Yongwei
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;; FILE REFERENCE: 38-21(535)B
;; CURRENT APPLICATION NUMBER: US/10/767,701
;; CURRENT FILING DATE: 2004-01-29
;; NUMBER OF SEQ ID NOS: 63128
;; SEQ ID NO 9911
;; LENGTH: 887
;; TYPE: DNA
;; ORGANISM: Sorghum bicolor
;; FEATURE:
;; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS40598_1
US-10-767-701-9911

Query Match
Best Local Similarity 7.5%; Score 34.2; DB 6; Length 887;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 203 TGGCGGTTTGGCAGAGAGTAGTACACCGGCAAGATTGACACAGAGATT 262
Db 433 TGTCAATGATTCCTAAAGTGATGATCAATTGGGGAATCTGATTGCTCCAGAG 492
Qy 263 ATAACTTGATATATGATCAGCGGCGGACATGCCAGTATTCGCAAGTG 322
Db 493 ATGGCTTAAATGTAAGCTTAAGACACAAACCAATGATACGAGTTCTAAGATA 552
Qy 323 CTATAGTAATAGTG 337
Db 553 GCAATGGGGAATG 567

RESULT 5
US-10-771-241-139
;; Sequence 139, Application US/10771241
;; GENERAL INFORMATION:
;; APPLICANT: Zyskind, Judith
;; APPLICANT: Forsyth, R. Allyn
;; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;; FILE REFERENCE: ELITRA.001C1
;; CURRENT APPLICATION NUMBER: US/10/771,241
;; CURRENT FILING DATE: 2004-02-03
;; PRIOR APPLICATION NUMBER: 09/492,709
;; PRIOR FILING DATE: 2000-01-27
;; PRIOR APPLICATION NUMBER: 60/117,405
;; PRIOR FILING DATE: 1999-01-27
;; NUMBER OF SEQ ID NOS: 485
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 139
;; LENGTH: 1149
;; TYPE: DNA
;; ORGANISM: E. coli
US-10-771-241-139

Query Match
Best Local Similarity 7.4%; Score 33.8; DB 6; Length 1149;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 259 GATTATACCTTGATATATGATCAGCGGCGGACATGCCAGATTCGCA 318
Db 442 GATATACCTGATATATGATCAGCGGCGGACATGCCAGATTCGCA 318
Qy 319 GGTCTTATAGTAATAGTGCATGATTAATTCAGAAAAGTTCTGATA 367
Db 502 GCACCGAGTGTACAGTTCTTATATTCGAGAGCTTATGATA 550

RESULT 6

[illegible]

PRIOR APPL.

RESULT 9
 US-10-417-884A-2325
 Sequence 2325, Application US/10417884A
 GENERAL INFORMATION:
 APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354
 COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/417,884A
 FILING DATE: 17-Apr-2003
 PRIORITY DATA:

```

APPLICATION NUMBER: US/09/107,533
FILING DATE: 30-Jun-1998
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Denae
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781) 893-5007
TELEFAX: (781) 893-8277
INFORMATION FOR SEQ ID NO: 2325:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 base pairs
TYPE: nucleic acid
STRANDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...525
SEQUENCE DESCRIPTION: SEQ ID NO: 2325:
US-10-417-884A-2325

```

Query Match	6.9%;	Score	31.4;	DB	6;	Length	525;
Best Local Similarity	54.9%;	Pred. No.	2	7.			

	Conservative	U	Mismatches	51	Indels	0	Gaps	0
QY	270	TGCATATATGATTCAGGCGGCGAGTCCCAAGATGCAGTATTTGCAAGTCTTATGG	329					
Db	213	TGAATATCTGATTAAGGCGAAGAAGACAGTGGCCATATTTCTCAGGGATTAAGTA	272					
QY	330	TAAATCTGAGATGATTAATCCAGAAAGGTTCTGGTAATTAAGCAAAATATACAC	382					
Db	273	CTATTTGATGATGAGTGTGAAATCATATTTGGGATTAAGACAAATATGTAC	325					

```

RESULT 10
US-60-546-745-36
; Sequence 36. Application US/60546745
; GENERAL INFORMATION:
; APPLICANT: Kiaenhamer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Alterman, Eric
; APPLICANT: Cano, Raul
; APPLICANT: Hamrick, Alice
; TITLE OF INVENTION: LACTOBACILLUS ACIDOPHILUS NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: PROPEASE HOMOLOGUES AND USES THEREOF
; FILE REFERENCE: 5051,604PR2
; CURRENT APPLICATION NUMBER: US/60/546,745
; CURRENT FILING DATE: 2004-02-23
; NUMBER OF SEQ ID NOS: 96
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 36
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Lactobacillus acidophilus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1245)
; OTHER INFORMATION: pepT amino tripeptidase T (EC 3.4.11.-) ORF# 1190
; US-60-546-745-36

```

Query March	6.9%	Score 31.4;	DB 7;	Length 1245;
Best Local Similarity	49.7%;	Pred. No. 3.6;		
Matches	80;	Conservative	0;	Mismatches 81;
				Indels 0;
				Gaps 0;

QY 240 AAGATTGACACAGACGAGATTATPACCTTGATATATTTGATCAGCGCGGACGTGCCAA 239

Dd 129 AGATCTTGAGGAATTTAGGCGCTTGAAAGAAATTCACACATCAAAAGCGTGTATGTATAT 188

QY 300 CAAAGCCAGATTTGCGCAAGGTCCTTATGGTAATPCTGCATGATTTATCCAGAAAGGTTTC 359

Dd 189 CGCTGAAATTCCTTCAAAATGTGATTAAGATGTAACCTGTATGGATTTCTTGCGCAATAG 248

QY 360 TGGTAAATAAGCAAAATATTTACACAGTATGTGATCAACCAAAAA 400

Dd 249 TGAATACAGTGATTCATTCAGAAAATGTTAAACCAACA 289

RESULT 11
US-10-786-065-3

```

; Sequence 3: Application US/10786065
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01098DIV 11
; CURRENT APPLICATION NUMBER: US/10/786,065
; CURRENT FILING DATE: 2004-02-26
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 53332
; TYPE: DNA
; ORGANISM: Homo sapiens
; GS-10-786-065-3

```

Query Match	6.9%	Score 31.4;	DB 6;	Length 53332;
Best Local Similarity	61.7%	Pred NO	13.	

Accession	Conserved	U	Mismatches	Indels	Gaps
QY	336	TGCAGTATTATCCAGAAAGTTCTGTGTAATAAAGCAATATTACACAGATGATGACCA	395		
DbD	8632	TGTATGAGTATTCACACAAATATGTATATTAAGCACTTACACAGACTAGTTTTCA	8693		
QY	396	AAAAACGGCAATTGTAGTGCA	416		
DbD	8692	GAGTTCAACAAATGTTGACCA	8712		

```

RESULT 12
US-10-767-471--8836/C
; Sequence 8836, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C1001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8836
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471--8836

```

Query Match	6.8%;	Score 30.8;	DB 6;	Length 201;
Best Local Similarity	54.6%;	Pred No 3;		

326 ATGCTAACTACGCGATGATTTATCCAGAAAGCTTCTGTAATATAAGCAAAATATTACACAT 385
 142 ATGGGAGATGGGAGCGAGAAATCCAGATGGTGATGATTAATRGGAGGCGATTTTCCACAG 83
 386 ATGCTACTCAAAAACCGCAATTGATGTCAGAGACAGTCGCAATG 433

Db 82 AGATTATCATGTACATGATGAGCAGACAGAAACAAAGCTGG 35

RESULT 13

US-10-767-471-34493/c
; Sequence 34493, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34493
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471-34493

Query Match

Best Local Similarity 54.6%; Score 30.8; DB 6; Length 201;
Matches 59; Conservative 1; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ATGGTAATACCTGCGATGATTATCCGAAAGTCTCGTAATTAAGCAATATTACACAGT 385

Db 142 ATGGAGATGAGGCGGCAATCCAAATGCGATGATGATTAATTAAGAGGCTATTTCACAG 83

Qy 386 ATGGTACTCAAAAAACGCAATTGTAGTCAGACAGACAGTCGCAATG 433

Db 82 AGATTATCATGTACATGATGAGCAGACAGAAACAAAGCTGG 35

RESULT 14

US-10-767-471-501/c
; Sequence 501, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471-501

Query Match

Best Local Similarity 54.6%; Score 30.8; DB 6; Length 1227;
Matches 59; Conservative 1; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ATGGTAATACCTGCGATGATTATCCGAAAGTCTCGTAATTAAGCAATATTACACAGT 385

Db 476 ATGGAGATGAGGCGGCAATCCAAATGCGATGATGATTAATTAAGAGGCTATTTCACAG 417

Qy 386 ATGGTACTCAAAAAACGCAATTGTAGTCAGACAGACAGTCGCAATG 433

Db 416 AGATTATCATGTACATGATGAGCAGACAGAAACAAAGCTGG 369

RESULT 15

US-10-767-471-10718/c
; Sequence 10718, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01505

; CURRENT APPLICATION NUMBER: US/10/767,471

; CURRENT FILING DATE: 2004-01-30

; NUMBER OF SEQ ID NOS: 50231

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10718

; LENGTH: 76573

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)...(76573)

; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-

US-10-767-471-10718

Query Match

Best Local Similarity 54.6%; Score 30.8; DB 6; Length 76573;
Matches 59; Conservative 1; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ATGGTAATACCTGCGATGATTATCCGAAAGTCTCGTAATTAAGCAATATTACACAGT 385

Db 69821 ATGGAGATGAGGCGGCAATCCAAATGCGATGATGATTAATTAAGAGGCTATTTCACAG 69762

Qy 386 ATGGTACTCAAAAAACGCAATTGTAGTCAGACAGACAGTCGCAATG 433

Db 69761 AGATTATCATGTACATGATGAGCAGACAGAAACAAAGCTGG 69714

Search completed: March 17, 2004, 08:25:37
Job time: 72.6301 secs.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 307.572 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48

Sequence: 1 tatgatcgcgtcgtaccgcg.....cccatgaatgacacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending Patents NA Main:

```

1: /cgn2_6/ptodata/2/pna/US0108 COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06 COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07 COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08 COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09 COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10 COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US11 COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US12 COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US13 COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US14 COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US15 COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US16 COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US17 COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US18 COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US19 COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US20 COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US21 COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US22 COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US23 COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US24 COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US25 COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US26 COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US27 COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US28 COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US29 COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US30 COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US31 COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US32 COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US33 COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US34 COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US35 COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US36 COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US37 COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US38 COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US39 COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US40 COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US41 COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US42 COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US43 COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US44 COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US45 COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US46 COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US47 COMB.seq.*

```

```

44: /cgn2_6/ptodata/2/pna/US1008 COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US1018 COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US1018 COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US1028 COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US1028 COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US1038 COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US1038 COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US1048 COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US1048 COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US1078 COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US1078 COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US1078 COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6000 COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6001 COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6002 COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6003 COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6004 COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6005 COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6006 COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6007 COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6008 COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6009 COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6010 COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6011 COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6012 COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6013 COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6014 COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6015 COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6016 COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6017 COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6018 COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6019 COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US6020 COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US6021 COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US6022 COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US6023 COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US6023 COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US6024 COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US6025 COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US6026 COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US6027 COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US6028 COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US6029 COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US6030 COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US6031 COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US6032 COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US6033 COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US6034 COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US6035 COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US6036 COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US6037 COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US6038 COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US6039 COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US6040 COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US6041 COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US6042 COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US6043 COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US6044 COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US6045 COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US6046 COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US6047 COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US6048 COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US6049 COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US6050 COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US6051 COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US6052 COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US6053 COMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	48	100.0	48	23	US-09-543-407-9	Sequence 9, Appl
C 2	48	100.0	78	23	US-09-543-407-52	Sequence 52, Appl
C 3	48	100.0	78	23	US-09-543-407-53	Sequence 53, Appl
C 4	48	100.0	456	23	US-09-543-407-11	Sequence 11, Appl
C 5	48	100.0	456	23	US-09-543-407-13	Sequence 13, Appl
C 6	48	100.0	456	23	US-09-543-407-15	Sequence 15, Appl
C 7	48	100.0	456	23	US-09-543-407-17	Sequence 17, Appl
C 8	48	100.0	456	23	US-09-543-407-19	Sequence 19, Appl
C 9	48	100.0	456	23	US-09-543-407-21	Sequence 21, Appl
C 10	48	100.0	456	23	US-09-543-407-23	Sequence 23, Appl
C 11	48	100.0	456	23	US-09-543-407-25	Sequence 25, Appl
C 12	48	100.0	456	23	US-09-543-407-27	Sequence 27, Appl
C 13	48	100.0	456	23	US-09-543-407-29	Sequence 29, Appl
C 14	40.6	84.6	78	23	US-09-543-407-48	Sequence 48, Appl
C 15	40.6	84.6	78	23	US-09-543-407-49	Sequence 49, Appl
C 16	29.4	61.2	3	23	US-07-965-673A-9	Sequence 9, Appl
C 17	27.8	57.9	2547	48	US-10-282-122A-36154	Sequence 36154, A
C 18	26.8	55.8	107223	21	US-09-528-237A-850	Sequence 850, A
C 19	24.6	51.3	1780	15	US-09-107-433-2456	Sequence 2456, A
C 20	24.6	51.3	2448	53	US-10-617-320-2456	Sequence 2456, A
C 21	24.6	51.3	2544	52	US-10-472-928-1519	Sequence 1519, A
C 22	24.6	51.3	2547	48	US-10-282-122A-37710	Sequence 37710, A
C 23	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 24	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 25	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 26	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 27	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 28	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 29	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 30	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 31	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 32	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 33	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 34	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 35	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 36	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 37	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 38	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 39	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 40	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 41	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 42	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 43	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 44	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl
C 45	24.6	51.3	2547	48	US-10-640-833-77	Sequence 77, Appl

ALIGNMENTS

RESULT 1
US-09-543-407-9
Sequence 9, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT FILING DATE: 2000-04-05
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 48
TYPE: DNA
ORGANISM: Leishmania major

US-09-543-407-9

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 48;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 TATGATCAGCTGGTTACCCGTTGTTACCATGAATGGACATGCA 48
1 TATGATCAGCTGGTTACCCGTTGTTACCATGAATGGACATGCA 48

RESULT 2

US-09-543-407-52/C
Sequence 52, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT FILING DATE: 2000-04-05
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 52
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 78;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-09-543-407-52

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 78;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 TATGATCAGCTGGTTACCCGTTGTTACCATGAATGGACATGCA 48
48 TATGATCAGCTGGTTACCCGTTGTTACCATGAATGGACATGCA 1

RESULT 3

US-09-543-407-53
Sequence 53, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT FILING DATE: 2000-04-05
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 78;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 TATGATCAGCTGGTTACCCGTTGTTACCATGAATGGACATGCA 48
1 TATGATCAGCTGGTTACCCGTTGTTACCATGAATGGACATGCA 48

RESULT 4
US-09-543-407-11; Sequence 11, Application US/09543407
; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-11

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 48
DB 382 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 429

RESULT 5
US-09-543-407-13; Sequence 13, Application US/09543407
; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-13

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 48
DB 352 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 399

RESULT 6
US-09-543-407-15

; Sequence 15, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-15

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 48
DB 67 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 114

RESULT 7
US-09-543-407-17; Sequence 17, Application US/09543407
; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 17

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-17

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 48
DB 127 TATGATCAGCTGCTTACCCGTTGTTACCATGAATGGACATGCA 174

RESULT 8
US-09-543-407-19; Sequence 19, Application US/09543407
; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543,407
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 19
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-19

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 48
196 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 243

RESULT 9
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 48
262 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 309

RESULT 10
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05

;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 23
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-23

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 48
331 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 378

RESULT 11
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-25

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
1 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 48
172 TATGATCAGCTGCTTACCCGCTGTTGTTACCATGAATGGACATGCA 219

RESULT 12
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 456
; TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-27

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 48
Db 241 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 288

RESULT 13
US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-29

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 48
Db 307 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 354

RESULT 14
US-09-543-407-48/c
Sequence 48, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 48
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer
US-09-543-407-48

Query Match 84.6%; Score 40.6; DB 23; Length 78;
Best Local Similarity 91.5%; Pred. No. 1.1e-05;
Matches 43; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 47
Db 48 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 2

RESULT 15
US-09-543-407-49
Sequence 49, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 49
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer
US-09-543-407-49

Query Match 84.6%; Score 40.6; DB 23; Length 78;
Best Local Similarity 91.5%; Pred. No. 1.1e-05;
Matches 43; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 47
Db 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 47

Search completed: March 18, 2004, 02:46:57
Job time : 311.872 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 7.32948 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48

Sequence: 1 tatgacgcgcgtgtaccgcg.....cccatgaatgcacatgca 48

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_New.*
1: /cgn2_6/ptodata/2/pna/PCR_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.2	48.3	435	US-10-767-795-1394	Sequence 1394, Ap
2	23.2	48.3	539	US-10-767-795-1392	Sequence 1392, Ap
3	23.2	48.3	545	US-10-767-701-29396	Sequence 29396, A
4	23.2	48.3	743	US-10-767-701-11493	Sequence 11493, A
5	22.2	46.2	1035	US-10-781-014-167	Sequence 167, App
6	22	45.8	801	US-10-779-543-5864	Sequence 5864, Ap
7	21.6	45.0	201	US-10-767-471-34311	Sequence 34311, A
8	21.6	45.0	2895	US-10-767-701-14458	Sequence 14458, A
9	21.6	45.0	546025	US-10-767-471-10717	Sequence 10717, A
10	21.4	44.6	607	US-10-767-701-31393	Sequence 31393, A
11	21	43.8	26377	US-10-765-790-41	Sequence 41, App1
12	20.8	43.3	175737	US-10-783-271-34	Sequence 34, App1
13	20.6	42.9	421	US-10-779-543-15761	Sequence 15761, A
14	20.6	42.9	905	US-10-781-469-49	Sequence 49, App1
15	20.6	42.9	198161	US-10-775-169-52	Sequence 52, App1
16	20.4	42.5	201	US-10-767-471-11899	Sequence 11899, A
17	20.4	42.5	1404	US-10-417-884A-2521	Sequence 2521, Ap
18	20.4	42.5	318752	US-10-767-471-10553	Sequence 10553, A
19	20.2	42.1	532	US-10-767-701-30765	Sequence 30765, A
20	20.2	42.1	1203	US-10-781-014-169	Sequence 169, App
21	20.2	42.1	1203	US-10-417-884A-2612	Sequence 2612, App
22	20.2	42.1	4503	US-10-417-375A-31	Sequence 31, App1
23	20.2	42.1	106776	US-10-417-375A-30	Sequence 30, App1
24	20	41.7	596	US-10-767-701-657	Sequence 657, App
25	20	41.7	746	US-10-779-543-5903	Sequence 5903, Ap
26	20	41.7	863	US-10-767-701-8642	Sequence 8642, Ap

27	20	41.7	3213	6	US-10-417-884A-62	Sequence 62, App1
28	20	41.7	36135	6	US-10-100-683-11827	Sequence 11827, A
29	19.8	41.2	390	6	US-10-779-543-12187	Sequence 12187, A
30	19.8	41.2	594	6	US-10-417-884A-380	Sequence 380, App
31	19.8	41.2	600	7	US-60-545-213-391	Sequence 391, App
32	19.8	41.2	600	7	US-60-545-213-392	Sequence 392, App
33	19.8	41.2	600	7	US-60-545-213-542	Sequence 542, App
34	19.8	41.2	600	7	US-60-545-213-4663	Sequence 4663, App
35	19.8	41.2	600	7	US-60-545-213-4664	Sequence 4664, App
36	19.8	41.2	600	7	US-60-545-213-4814	Sequence 4814, App
37	19.8	41.2	757	6	US-10-100-683-5410	Sequence 5410, App
38	19.8	41.2	844	6	US-10-767-701-10879	Sequence 10879, App
39	19.8	41.2	7652	1	PCT-US03-31941A-1	Sequence 1, App1
40	19.8	41.2	8684	1	PCT-US03-31941A-2	Sequence 2, App1
41	19.8	41.2	10011	1	PCT-US03-31941A-19	Sequence 19, App1
42	19.6	40.8	201	6	US-10-767-471-34488	Sequence 34488, A
43	19.6	40.8	201	6	US-10-767-471-34490	Sequence 34490, A
44	19.6	40.8	201	6	US-10-767-471-34556	Sequence 34556, A
45	19.6	40.8	1013	6	US-10-767-701-12512	Sequence 12512, A

ALIGNMENTS

RESULT 1	US-10-767-795-1394	Sequence 1394, Application US/10767795
GENERAL INFORMATION:		
APPLICANT:	Kovalic, David K.	
APPLICANT:	Caoy, Yongwei	
APPLICANT:	Zhou, Yihua	
TITLE OF INVENTION:	Nucleic Acid Molecules and Other Molecules Associated With	
FILE REFERENCE:	38-21(53534)B	
CURRENT FILING DATE:	2004-01-30	
NUMBER OF SEQ ID NOS:	117596	
SEQ ID NO 1394		
LENGTH:	435	
TYPE:	DNA	
ORGANISM:	Gossypium hirsutum	
FEATURE:		
OTHER INFORMATION:	Clone ID: UC-gsromu33B172b05b1	
US-10-767-795-1394		
Query Match	48.3%; Score 23.2; DB 6; Length 435;	
Best Local Similarity	70.5%; Pred. No. 1.6;	
Matches	31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;	
QY	3 TGATCAGCTGGTACCGGTGTTGCCCAAGAAATGCACATG 46	
DB	73 TGAATACCTCATTAACCTGTTGGTTACCAAGAAAGACATG 116	
RESULT 2	US-10-767-795-1392	
Sequence 1392, Application US/10767795		
GENERAL INFORMATION:		
APPLICANT:	Kovalic, David K.	
APPLICANT:	Caoy, Yongwei	
APPLICANT:	Zhou, Yihua	
TITLE OF INVENTION:	Nucleic Acid Molecules and Other Molecules Associated With	
FILE REFERENCE:	38-21(53534)B	
CURRENT FILING DATE:	2004-01-30	
NUMBER OF SEQ ID NOS:	117596	
SEQ ID NO 1392		
LENGTH:	539	
TYPE:	DNA	
ORGANISM:	Gossypium hirsutum	
FEATURE:		
OTHER INFORMATION:	Clone ID: UC-gsromu33B027f06b1	

US-10-767-795-1392

Query Match

Best Local Similarity 48.3%; Score 23.2; DB 6; Length 539;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;QY 3 TGATCAGCTGTTACCCGCTGTTGTTACCATGAATGACATG 46
DB 29 TGATTAATCATCACTCGTTGTTGTTACCAAGAGACATG 72

RESULT 3

US-10-767-701-29396/c

Sequence 29396, Application US/10767701

GENERAL INFORMATION:

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53535)B

CURRENT APPLICATION NUMBER: US/10/767,701

NUMBER OF SEQ ID NOS: 2004-01-29

SEQ ID NO 29396

LENGTH: 545

TYPE: DNA

ORGANISM: Sorghum bicolor

FEATURE:

OTHER INFORMATION: Clone ID: 9299890

US-10-767-701-29396

Query Match

Best Local Similarity 48.3%; Score 23.2; DB 6; Length 545;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;QY 5 ATCAGCTGTTACCCGCTGTTGTTACCATGAATG 40
DB 532 ATCAGCTGTTACCCGCTGTTGTTACCATGAATG 497

RESULT 4

US-10-767-701-11493

Sequence 11493, Application US/10767701

GENERAL INFORMATION:

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53535)B

CURRENT APPLICATION NUMBER: US/10/767,701

NUMBER OF SEQ ID NOS: 2004-01-29

SEQ ID NO 11493

LENGTH: 743

TYPE: DNA

ORGANISM: Sorghum bicolor

FEATURE:

OTHER INFORMATION: Clone ID: SORBI-28MAV03-CLUS45512_1

US-10-767-701-11493

Query Match

Best Local Similarity 48.3%; Score 23.2; DB 6; Length 743;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;QY 2 ATGATCAGCTGTTACCCGCTGTTGTTACCATGAATG 45
DB 269 ATGATCAGCTGTTACCCGCTGTTGTTACCATGAATG 312

RESULT 5

US-10-781-014-167

US-10-781-014-167

Sequence 167, Application US/10781014

GENERAL INFORMATION:

APPLICANT: Pompeius, Markus

APPLICANT: Kroger, Burkhard

APPLICANT: Schroder, Hartwig

APPLICANT: Zelder, Oskar

APPLICANT: Habenhauer, Gregor

TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS

INVOLVED IN CARBON METABOLISM AND ENERGY

FILE REFERENCE: BGI-126CPN

CURRENT APPLICATION NUMBER: US/10/781,014

NUMBER OF SEQ ID NOS: 2004-02-17

SEQ ID NO 167

LENGTH: 1035

TYPE: DNA

ORGANISM: Corynebacterium glutamicum

FEATURE:

NAME/KEY: CDS

LOCATION: (101) ..(1012)

OTHER INFORMATION: RXN02920

US-10-781-014-167

Query Match

Best Local Similarity 46.2%; Score 22.2; DB 6; Length 1035;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;QY 3 TGATCAGCTGTTACCCGCTGTTGTTACCATGAATG 45
DB 280 TGATCAGCTGTTACCCGCTGTTGTTACCATGAATG 322

RESULT 6

US-10-779-543-5864

Sequence 5864, Application US/10779543

GENERAL INFORMATION:

APPLICANT: Williams et al

TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED

FILE REFERENCE: 2300-21302

CURRENT APPLICATION NUMBER: US/10/779,543

NUMBER OF SEQ ID NOS: 2004-02-12

SEQ ID NO 167

LENGTH: 1035

TYPE: DNA

ORGANISM: Corynebacterium glutamicum

FEATURE:

NAME/KEY: CDS

LOCATION: (101) ..(1012)

OTHER INFORMATION: RXN02920

US-10-779-543-5864

Query Match

Best Local Similarity 46.2%; Score 22.2; DB 6; Length 1035;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;QY 3 TGATCAGCTGTTACCCGCTGTTGTTACCATGAATG 45
DB 280 TGATCAGCTGTTACCCGCTGTTGTTACCATGAATG 322

RESULT 6

US-10-779-543-5864

Sequence 5864, Application US/10779543

GENERAL INFORMATION:

APPLICANT: Williams et al

TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED

FILE REFERENCE: 2300-21302

CURRENT APPLICATION NUMBER: US/10/779,543

NUMBER OF SEQ ID NOS: 2004-02-12

SEQ ID NO 167

LENGTH: 1035

TYPE: DNA

ORGANISM: Corynebacterium glutamicum

FEATURE:

NAME/KEY: CDS

LOCATION: (101) ..(1012)

OTHER INFORMATION: RXN02920

US-10-779-543-5864

Oy 1 TATGATCAGCTGGTTACCCGTGTGTTTACCACATGAATAAGCACAA 44
 ||||| | | | | | | | | | |
Db 155 TTGGATCAGTTTCATATTCTCGTGCGGTACTPATGAAGAAGAACAA 198

```

RESULT 10
US-10-767-701-31393/c
; Sequence 31393, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5335)B

```

```

; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 31393
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(607)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: 18067944
; US-10-767-701-31393

Query Match      44.6%; Score 21.4; DB 6; Length 607;
Best Local Similarity 66.0%; Pred. No. 9.5;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY      2 ATGATCAGCTGTTACCGGTGTTGTTACCATGAATGCAATGCA 48
Db      306 AGATGCGCTGTGAGCTTTATTATGACACCATTAATGCAACA 260

RESULT 11
US-10-765-790-41/c
; Sequence 41, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 41
; LENGTH: 26371
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-765-790-41

Query Match      43.8%; Score 21; DB 6; Length 26371;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY      2 ATGATCAGCTGTTACCGGTGTTGTTACCATGAATGCAATG 46
Db      15911 AAGCCAGCTGTGATCTTGATGTCTCCCAACAATAGAATG 15867

RESULT 12
US-10-783-271-34
; Sequence 34, Application US/10783271
; GENERAL INFORMATION:
; APPLICANT: Veridex, LLC
; APPLICANT: Wang, Yixin
; TITLE OF INVENTION: BREAST CANCER PROGNOSTICS
; FILE REFERENCE: VDX-5003 USNP
; CURRENT APPLICATION NUMBER: US/10/783,271
; CURRENT FILING DATE: 2004-02-20
; NUMBER OF SEQ ID NOS: 111
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 175737
; TYPE: DNA
```

```

; ORGANISM: human
; US-10-783-271-34

Query Match      43.3%; Score 20.8; DB 6; Length 175737;
Best Local Similarity 70.0%; Pred. No. 58;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY      9 GGTGTTACCGGTGTTGTTACCATGAATGCAATGCA 48
Db      116051 GCTGCTCATCTGGAGCTTCCCTCGAATGCACTCCA 116090

RESULT 13
US-10-779-543-15761/c
; Sequence 15761, Application US/10779543
; GENERAL INFORMATION:
; APPLICANT: Williams et al
; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
; FILE REFERENCE: 2300-21302
; CURRENT APPLICATION NUMBER: US/10/779,543
; CURRENT FILING DATE: 2004-02-12
; PRIOR APPLICATION NUMBER: 10/076,555
; PRIOR FILING DATE: 2002-02-15
; PRIOR APPLICATION NUMBER: 09/217,471
; PRIOR FILING DATE: 1998-12-21
; PRIOR APPLICATION NUMBER: 60/068,755
; PRIOR FILING DATE: 1997-12-23
; PRIOR APPLICATION NUMBER: 60/080,664
; PRIOR FILING DATE: 1998-04-03
; PRIOR APPLICATION NUMBER: 60/105,234
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15761
; LENGTH: 421
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-779-543-15761

Query Match      42.9%; Score 20.6; DB 6; Length 421;
Best Local Similarity 74.3%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      11 TGGTTACCGGTGTTGTTACCATGAATGCAATG 45
Db      225 TTGTTACTGTGCGATTTACCATGAATATTCACAT 191

RESULT 14
US-10-781-469-49
; Sequence 49, Application US/10781469
; GENERAL INFORMATION:
; APPLICANT: Reddy, Koopa
; APPLICANT: Guegler, Karl, J.
; APPLICANT: Au-Young, Janice K.
; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING RECEPTOR PROTEINS
; FILE REFERENCE: PA-0014-1 CON
; CURRENT APPLICATION NUMBER: US/10/781,469
; CURRENT FILING DATE: 2004-02-17
; PRIOR APPLICATION NUMBER: US 09/496,321
; PRIOR FILING DATE: 2000-02-01
```


PRIOR APPLICATION NUMBER: US 60/118,905
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PERL Program
SEQ ID NO 49
LENGTH: 905
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID NO: 2832568
US-10-781-469-49

Query Match 42.9%; Score 20.6; DB 6; Length 905;
Best Local Similarity 67.4%; Pred. No. 22;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTACCCGTTGTTACCATGAATGGACACA 44
Db 605 ATAACTGGGTCCACGAGTCATAGCCCATGAATGGCCCA 647

RESULT 15
US-10-775-169-52
Sequence 52, Application US/10775169
GENERAL INFORMATION:
APPLICANT: Wyeth
APPLICANT: Butczynski, Michael
APPLICANT: Twine, Natalie
APPLICANT: Dornier, Andrew
APPLICANT: Trepicchio, William
TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
FILE REFERENCE: AM101080 (031896-013000)
CURRENT APPLICATION NUMBER: US/10/775,169
CURRENT FILING DATE: 2004-02-11
NUMBER OF SEQ ID NOS: 5278
SOFTWARE: PatentIn version 3.2
SEQ ID NO 52
LENGTH: 198161
TYPE: DNA
ORGANISM: Homo sapiens
US-10-775-169-52

Query Match 42.9%; Score 20.6; DB 6; Length 198161;
Best Local Similarity 67.4%; Pred. No. 72;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 6 TCAGCTGTACCCGTTGTTACCATGAATGGACATGCA 48
Db 175379 TCACATGTGTCTCATGTTCTTAACACGAATGTCTCAGGGA 175421

Search completed: March 17, 2004, 08:25:38
Job time : 8.32948 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456
Sequence: 1 atgaacttttaaaagtgc.....ccacgctaacagctactaa 456

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents NA Main:*

1: /cgn2_6/ptodata/2/pna/US0954_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US0957_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US0958_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US0959_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US0960_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US0961_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US0962_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US0963_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US0964_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US0965_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US0966_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US0967_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US0968_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US0969_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US0970_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US0971_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US0972_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US0973_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US0974_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US0975_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US0976_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US0977_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US0979_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US0980_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US0981_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US0982_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US0983_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US0984_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US0985_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US0986_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US0987_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US0988_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US0989_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US0990_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US0991_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US0992_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US0993_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US0994_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US0995_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US0996_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US0997_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US1008_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US1009_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US1010_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US1011_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US1012_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US1013_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US1014_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US1015_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US1016_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US1017_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US1018_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US1019_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US1020_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US1021_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US1022_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US1023_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US1024_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US1025_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US1026_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US1027_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US1028_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US1029_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US1030_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US1031_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US1032_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US1033_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US1034_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US1035_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US1036_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US1037_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US1038_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US1039_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US1040_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US1041_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US1042_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US1043_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US1044_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US1045_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US1046_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US1047_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US1048_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US1049_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US1050_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US1051_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US1052_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US1053_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US1054_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US1055_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US1056_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US1057_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US1058_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US1059_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US1060_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US1061_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US1062_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US1063_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US1064_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US1065_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US1066_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US1067_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US1068_COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US1069_COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US1070_COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US1071_COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US1072_COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US1073_COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US1074_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	23	US-09-543-407-11
2	404.8	88.8	456	6	US-08-233-642A-56
3	404.8	88.8	456	23	US-09-543-407-11
4	376	82.5	456	23	US-09-543-407-13
5	361.6	79.3	456	23	US-09-543-407-19
6	355.2	77.3	456	23	US-09-543-407-17
7	353.6	77.5	456	23	US-09-543-407-23
8	353.6	77.5	456	23	US-09-543-407-25
9	352	77.2	456	23	US-09-543-407-21
10	350.4	76.8	456	23	US-09-543-407-27
11	350.4	76.8	456	23	US-09-543-407-29
12	347.2	76.1	456	23	US-09-543-407-29
13	315.2	69.1	456	23	US-08-233-642A-54
14	222.4	48.8	477	17	US-09-252-691C-172
15	222.4	48.8	477	17	US-09-252-691C-172
16	222.4	48.8	477	51	US-10-417-886-172
17	216	47.4	456	23	US-09-543-407-3
18	216	47.4	456	13	US-08-978-878-3
19	216	47.4	456	31	US-09-741-873B-1
20	213.6	46.8	453	92	US-60-352-946-1
21	213.6	46.8	453	101	US-60-444-371-1
22	172	37.7	396	13	US-08-978-878-1
23	163.2	35.8	360	31	US-09-741-873B-1
24	52.2	11.4	78	23	US-09-543-407-53
25	49.4	10.8	78	23	US-09-543-407-52
26	48	10.5	78	23	US-09-543-407-9
27	40.8	8.9	78	23	US-09-543-407-48
28	40.8	8.9	100	45	US-10-146-492B-79
29	40.6	8.9	78	23	US-09-543-407-49
30	39.4	8.6	2000	36	US-09-887-272A-5263
31	39	8.6	100	45	US-10-146-492B-78
32	37.4	8.2	561	17	US-09-252-691C-171
33	37.4	8.2	561	17	US-09-252-691C-171
34	37.4	8.2	561	51	US-10-417-886-171
35	37.4	8.2	1304	51	US-10-437-963-19773
36	35.4	7.8	100	45	US-10-146-492B-77
37	35.4	7.8	3411	1	PCT-US01-08631-24330
38	35.4	7.8	3412	1	PCT-US01-08631-24948
39	34.6	7.6	456	23	US-09-543-407-4
40	34.2	7.5	385	18	US-09-333-535A-288
41	34.2	7.5	385	25	US-09-615-606A-909
42	34.2	7.5	396	64	US-60-089-524-4522
43	34.2	7.5	616	51	US-10-424-599-88945
44	34.2	7.5	880	27	US-09-654-617-2638
45	34.2	7.5	880	29	US-09-684-016-2638

ALIGNMENTS

RESULT 1
 US-09-543-407-11
 Sequence 11, Application US/09543407
 GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collins, S. Karen
 APPLICANT: Kay, William W.
 TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
 FILE REFERENCE: 920043.406
 PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 CURRENT APPLICATION NUMBER: US/09/543.407
 CURRENT FILING DATE: 2000-04-05
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 11
 LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence

FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 OTHER INFORMATION: encoding P13 from GPe3 of *Leishmania* major.
 US-09-543-407-11

Query Match 100.0%; Score 456; DB 23; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.6e-136;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAATGCGACATTCGACGATGTTCTGCGCTGCTGCTGCT 60
 1 ATGAACCTTTAAATGCGACATTCGACGATGTTCTGCGCTGCTGCTGCT 60
 QY 61 GGGGTCTTCAACATGCGGCGCGCGGATTCATATACGCGCGCAATGTTCCGCG 120
 61 GGGGTCTTCAACATGCGGCGCGCGGATTCATATACGCGCGCAATGTTCCGCG 120
 QY 121 CCGGACTCAACGTTGACATTTATCATGACGTTCCGCTGACGCTGCTGCTGCA 180
 121 CCGGACTCAACGTTGACATTTATCATGACGTTCCGCTGACGCTGCTGCTGCA 180
 QY 181 AGGATGCCCGTAATCTGAAACGACATTCACGAGCGGTTATGTTACGCGCGCAT 240
 181 AGGATGCCCGTAATCTGAAACGACATTCACGAGCGGTTATGTTACGCGCGCAT 240
 QY 241 GTAGGCGAGGTCGCGATATGATCTATTTGATGACGATGATGTTGAGATATAT 300
 241 GTAGGCGAGGTCGCGATATGATCTATTTGATGACGATGATGTTGAGATATAT 300
 QY 301 GGCACCATGACCATGACGATGACGATGACGATGACGATGACGATGACGATGAC 360
 301 GGCACCATGACCATGACGATGACGATGACGATGACGATGACGATGACGATGAC 360
 QY 361 AATAAGCGCGCTGATATATGATGATGATGATGATGATGATGATGATGATGAT 420
 361 AATAAGCGCGCTGATATATGATGATGATGATGATGATGATGATGATGATGAT 420
 QY 421 GGCATGCAACACGCGGCTTACGATGATGATGATGATGATGATGATGATGAT 480
 421 GGCATGCAACACGCGGCTTACGATGATGATGATGATGATGATGATGATGAT 480
 QY 481 GGCATGCAACACGCGGCTTACGATGATGATGATGATGATGATGATGATGAT 540
 481 GGCATGCAACACGCGGCTTACGATGATGATGATGATGATGATGATGATGAT 540

RESULT 2
 US-08-233-642A-56
 Sequence 56, Application US/08233642A
 GENERAL INFORMATION:
 APPLICANT: Kay, William W.
 APPLICANT: Collins, S. Karen
 APPLICANT: Clouthier, Sharon C.
 APPLICANT: Doran, James L.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
 NUMBER OF SEQUENCES: 58
 NUMBER OF SEQUENCES: 58
 CORRESPONDENCE ADDRESS:
 ADDRESS: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233.642A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 88.8%; Score 404.8; DB 6; Length 456;
Best Local Similarity 93.0%; Pred. No. 2,6e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
QY 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
DB 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
QY 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
QY 241 GTAGCCAGGGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 300
DB 241 GTAGCCAGGGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 300
QY 301 GCCACATCGACGAGTGAACGCTAAACCTCGATATTAATGATTAATGATTAATGATTAAT 360
DB 301 GCCACATCGACGAGTGAACGCTAAACCTCGATATTAATGATTAATGATTAATGATTAAT 360
QY 361 AATAACCGCGCGGTGTTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 420
DB 361 AATAACCGCGCGGTGTTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 420
QY 421 GCACATGCAACCAACGCGCGCTAACCAATATTA 456
DB 421 GGTTTGGCAACCAACGCGCGCTAACCAATATTA 456

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1
Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 2,6e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
QY 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
DB 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
QY 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
QY 241 GTAGCCAGGGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 300
DB 241 GTAGCCAGGGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 300
QY 301 GCCACATCGACGAGTGAACGCTAAACCTCGATATTAATGATTAATGATTAATGATTAAT 360
DB 301 GCCACATCGACGAGTGAACGCTAAACCTCGATATTAATGATTAATGATTAATGATTAAT 360
QY 361 AATAACCGCGCGGTGTTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 420
DB 361 AATAACCGCGCGGTGTTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 420
QY 421 GCACATGCAACCAACGCGCGCTAACCAATATTA 456
DB 421 GGTTTGGCAACCAACGCGCGCTAACCAATATTA 456

RESULT 4
US-09-543-407-13
Sequence 13, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding p73 from GPe3 of Leishmania major.

US-09-543-407-13
Query Match 82.5%; Score 376; DB 23; Length 456;
Best Local Similarity 89.0%; Pred. No. 5,1e-110;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGGCAGTGTCTGGCT 60
QY 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120

```
Db 61 GCGCGTGTCCACATGGGGCGGCGGTAATCTAATACGGCGGCAATAGTTCGGGC 120
Qy 121 CCGGACTCAAGCTTGAGCAATTTATGATGAGTTCGGTAAAGCTGCTGCTGCA 180
Db 121 CCGGACTCAAGCTTGAGCAATTTATGAGTTCGGTAAAGCTGCTGCTGCA 180
Qy 181 AGCGATGCCGCTTAATCTGAAACGACCTTCCCGAGACGGTATGTAAGCGCGCAT 240
Db 181 AGCGATGCCGCTTAATCTGAAACGACCTTCCCGAGACGGTATGTAAGCGCGCAT 240
Qy 241 GTAGGCCAGGGTGCAGATTAATGATCTATTGACTGACTGAGAAATGTTTCAAAATAT 300
Db 241 GTAGGCCAGGGTGCAGATTAATGATCTATTGACTGACTGAGAAATGTTTCAAAATAT 300
Qy 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATATTACTGTGCGCAATAGCAGGT 360
Db 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATATTACTGTGCGCAATAGCAGGT 360
Qy 361 AATAACGCGCGCTGTTAATTAATGATGAGTTCGCTTACCGCTTGTACCATGAATG 420
Db 361 CTGCTTACCGCTTGTACCATGAATGAGCAATGCAATGCAATGCTGCTGCTGCT 420
Qy 421 GCACATGCAACCAAGCCGACGCTTAACCGATTTAA 456
Db 421 GGTTTGGCAACCAAGCCGACGCTTAACCGATTTAA 456
```

RESULT 5

```
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Dorian, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19
```

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.3e-105;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```
Qy 1 ATGAAACTTTTAAAGTGGAGCATTCGACGCAATGTAATTCGCGAGTCTGGCT 60
Db 1 ATGAAACTTTTAAAGTGGAGCATTCGACGCAATGTAATTCGCGAGTCTGGCT 60
Qy 61 GGGCGTCTTCACAAATGGGCGGCGGCTAATCATTAACGGCGGCAATAGTTCGGC 120
Db 61 GGGCGTCTTCACAAATGGGCGGCGGCTAATCATTAACGGCGGCAATAGTTCGGC 120
Qy 121 CCGGACTCAAGCTTGAGCAATTTATGATGAGTTCGGTAAAGCTGCTGCTGCA 180
Db 121 CCGGACTCAAGCTTGAGCAATTTATGATGAGTTCGGTAAAGCTGCTGCTGCA 180
Qy 181 AGCGATGCCGCTTAATCTGAAACGACCTTCCCGAGACGGTATGTAAGCGCGCAT 240
Db 181 AGCGATGCCGCTTAATCTGAAACGACCTTCCCGAGACGGTATGTAAGCGCGCAT 240
```

```
Qy 241 GTAGGCCAGGGTGCAGATTAATGATCTATTGACTGACTGAGAAATGTTTCAAAATAT 300
Db 241 GTAGGCCAGGGTGCAGATTAATGATCTATTGACTGACTGAGAAATGTTTCAAAATAT 300
Qy 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATATTACTGTGCGCAATAGCAGGT 360
Db 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATATTACTGTGCGCAATAGCAGGT 360
Qy 361 AATAACGCGCGCTGTTAATTAATGATGAGTTCGCTTACCGCTTGTACCATGAATG 420
Db 361 AATAACGCGCGCTGTTAATTAATGATGAGTTCGCTTACCGCTTGTACCATGAATG 420
Qy 421 GCACATGCAACCAAGCCGACGCTTAACCGATTTAA 456
Db 421 GGTTTGGCAACCAAGCCGACGCTTAACCGATTTAA 456
```

RESULT 6

```
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Dorian, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17
```

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.6e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```
Qy 1 ATGAAACTTTTAAAGTGGAGCATTCGACGCAATGTAATTCGCGAGTCTGGCT 60
Db 1 ATGAAACTTTTAAAGTGGAGCATTCGACGCAATGTAATTCGCGAGTCTGGCT 60
Qy 61 GGGCGTCTTCACAAATGGGCGGCGGCTAATCATTAACGGCGGCAATAGTTCGGC 120
Db 61 GGGCGTCTTCACAAATGGGCGGCGGCTAATCATTAACGGCGGCAATAGTTCGGC 120
Qy 121 CCGGACTCAAGCTTGAGCAATTTATGATGAGTTCGGTAAAGCTGCTGCTGCA 180
Db 121 CCGGACTCAAGCTTGAGCAATTTATGATGAGTTCGGTAAAGCTGCTGCTGCA 180
Qy 181 AGCGATGCCGCTTAATCTGAAACGACCTTCCCGAGACGGTATGTAAGCGCGCAT 240
Db 181 AGCGATGCCGCTTAATCTGAAACGACCTTCCCGAGACGGTATGTAAGCGCGCAT 240
Qy 241 GTAGGCCAGGGTGCAGATTAATGATCTATTGACTGACTGAGAAATGTTTCAAAATAT 300
Db 241 GTAGGCCAGGGTGCAGATTAATGATCTATTGACTGACTGAGAAATGTTTCAAAATAT 300
Qy 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATATTACTGTGCGCAATAGCAGGT 360
Db 301 GCCACATCGACAGTGGAAAGCTTAAACTCCGATATTACTGTGCGCAATAGCAGGT 360
Qy 361 AATAACGCGCGCTGTTAATTAATGATGAGTTCGCTTACCGCTTGTACCATGAATG 420
Db 361 AATAACGCGCGCTGTTAATTAATGATGAGTTCGCTTACCGCTTGTACCATGAATG 420
```

Qy 421 GCACATGCAACCAACGGCGCTAACAGATTAA 456
Db 421 GCTTTGGCAACCAACGGCGCTAACAGATTAA 456

RESULT 7

US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-23

Query Match 77.5%; Score 353.6; DB 23; Length 456;

Best Local Similarity 86.0%; Pred. No. 8.6e-103; Indels 0; Gaps 0;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACGAATCGTAGTTCTGCGAGTCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACGAATCGTAGTTCTGCGAGTCTGGCT 60
Qy 61 GGGCTGTTCCACATGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
Db 61 GGGCTGTTCCACATGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGCTGAGATTATCAGTACGGTCCGCTACCGTGGCTGCTGCA 180
Db 121 CCGGACTCAAGCTGAGATTATCAGTACGGTCCGCTACCGTGGCTGCTGCA 180
Qy 181 AGCGATCGCGTAAATCTGAAGACCATTCACGAGCGGTATAGTAAACGGCGCAT 240
Db 181 AGCGATCGCGTAAATCTGAAGACCATTCACGAGCGGTATAGTAAACGGCGCAT 240
Qy 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAATGAACTGACTCAGAAATGTTT 300
Db 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAATGAACTGACTCAGAAATGTTT 300
Qy 301 GCCACCATCGACAGTGAAGCTAAATCCGATATTAATGTCGCGCAATAGCGCGT 360
Db 301 GCCACCATCGACAGTGAAGCTAAATCCGATATTAATGTCGCGCAATAGCGCGT 360
Qy 361 AATAAGCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTGTTAATCCATGAATG 420
Db 361 AATAAGCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTGTTAATCCATGAATG 420
Qy 421 GCACATGCAACCAACGGCGCTAACAGATTAA 456
Db 421 GCTTTGGCAACCAACGGCGCTAACAGATTAA 456

RESULT 8

US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-25

Query Match 77.5%; Score 353.6; DB 23; Length 456;

Best Local Similarity 86.0%; Pred. No. 8.6e-103; Indels 0; Gaps 0;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACGAATCGTAGTTCTGCGAGTCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACGAATCGTAGTTCTGCGAGTCTGGCT 60
Qy 61 GGGCTGTTCCACATGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
Db 61 GGGCTGTTCCACATGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGCTGAGATTATCAGTACGGTCCGCTACCGTGGCTGCTGCA 180
Db 121 CCGGACTCAAGCTGAGATTATCAGTACGGTCCGCTACCGTGGCTGCTGCA 180
Qy 181 AGCGATCGCGTAAATCTGAAGACCATTCACGAGCGGTATAGTAAACGGCGCAT 240
Db 181 AGCGATCGCGTAAATCTGAAGACCATTCACGAGCGGTATAGTAAACGGCGCAT 240
Qy 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAATGAACTGACTCAGAAATGTTT 300
Db 241 GTAGCGCAGGGTGGGATTAATAGTACTTGAATGAACTGACTCAGAAATGTTT 300
Qy 301 GCCACCATCGACAGTGAAGCTAAATCCGATATTAATGTCGCGCAATAGCGCGT 360
Db 301 GCCACCATCGACAGTGAAGCTAAATCCGATATTAATGTCGCGCAATAGCGCGT 360
Qy 361 AATAAGCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTGTTAATCCATGAATG 420
Db 361 AATAAGCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTGTTAATCCATGAATG 420
Qy 421 GCACATGCAACCAACGGCGCTAACAGATTAA 456
Db 421 GCTTTGGCAACCAACGGCGCTAACAGATTAA 456

RESULT 9

US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.8e-102;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
OY 1 ATGAACCTTTAAAGTGGAGCATTCGACGAACTGATGTTCTGCGAGTCTGGCT 60
Db 1 ATGAACCTTTAAAGTGGAGCATTCGACGAACTGATGTTCTGCGAGTCTGGCT 60
OY 61 GCGCTGTTCCACATGAGGCGCGCGGTAATCATACGGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCACATGAGTGTGTAACCGGTGTTGTTACCCATGGAATGGCAGATCCGGC 120
OY 121 CCGGACTCAACGTTGAGCATTTATCATGAGGTTCCGTAACGCTGCGCTTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCATGAGGTTCCGTAACGCTGCGCTTCTGCAA 180
OY 181 AGCGATGCCCGTAATCTGAAACGACCACTTACCGAGCGGTTATGTTAAGCGCGCGAT 240
Db 181 AGCGATGCCCGTAATCTGAAACGACCACTTACCGAGCGGTTATGTTAAGCGCGCGAT 240
OY 241 GTAGGCCAGGTTGCGGATTAATGATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGGCCAGGTTGCGGATTAATGATGTAATGTAATGTAATGTAATGTAATGTAAT 300
OY 301 GCCACATCGACCGGATGGAAGGCTTAAACCTCCGATTTACTGTCGCGCAATAGCGCGT 360
Db 301 GCCACATCGACCGGATGGAAGGCTTAAACCTCCGATTTACTGTCGCGCAATAGCGCGT 360
OY 361 AATAAGCCCGCTGTTAATTAATGATCAGCTGGTACCGGTGTTGTTACCCATGAATG 420
Db 361 AATAAGCCCGCTGTTAATTAATGATCAGCTGGTACCGGTGTTGTTACCCATGAATG 420
OY 421 GCACATGCAACGACCGCGGCTTAACGATTTAA 456
Db 421 GGTTTGCAACGACCGCGGCTTAACGATTTAA 456
```

RESULT 10
US-09-543-407-21
Sequence 21, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

Query Match 76.8%; Score 350.4; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 9.3e-102;

Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
OY 1 ATGAACCTTTAAAGTGGAGCATTCGACGAACTGATGTTCTGCGAGTCTGGCT 60
Db 1 ATGAACCTTTAAAGTGGAGCATTCGACGAACTGATGTTCTGCGAGTCTGGCT 60
OY 61 GCGCTGTTCCACATGAGGCGCGCGGTAATCATACGGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCACATGAGGCGCGCGGTAATCATACGGCGGCAATAGTTCCGGC 120
OY 121 CCGGACTCAACGTTGAGCATTTATCATGAGGTTCCGTAACGCTGCGCTTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCATGAGGTTCCGTAACGCTGCGCTTCTGCAA 180
OY 181 AGCGATGCCCGTAATCTGAAACGACCACTTACCGAGCGGTTATGTTAAGCGCGAT 240
Db 181 AGCGATGCCCGTAATCTGAAACGACCACTTACCGAGCGGTTATGTTAAGCGCGAT 240
OY 241 GTAGGCCAGGTTGCGGATTAATGATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGGCCAGGTTGCGGATTAATGATGTAATGTAATGTAATGTAATGTAATGTAAT 300
OY 301 GCCACATCGACCGGATGGAAGGCTTAAACCTCCGATTTACTGTCGCGCAATAGCGCGT 360
Db 301 GCCACATCGACCGGATGGAAGGCTTAAACCTCCGATTTACTGTCGCGCAATAGCGCGT 360
OY 361 AATAAGCCCGCTGTTAATTAATGATCAGCTGGTACCGGTGTTGTTACCCATGAATG 420
Db 361 AATAAGCCCGCTGTTAATTAATGATCAGCTGGTACCGGTGTTGTTACCCATGAATG 420
OY 421 GCACATGCAACGACCGCGGCTTAACGATTTAA 456
Db 421 GGTTTGCAACGACCGCGGCTTAACGATTTAA 456
```

RESULT 11
US-09-543-407-27

Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

Query Match 76.8%; Score 350.4; DB 23; Length 456;
Best Local Similarity 85.5%; Pred. No. 9.3e-102;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
OY 1 ATGAACCTTTAAAGTGGAGCATTCGACGAACTGATGTTCTGCGAGTCTGGCT 60
Db 1 ATGAACCTTTAAAGTGGAGCATTCGACGAACTGATGTTCTGCGAGTCTGGCT 60
OY 61 GCGCTGTTCCACATGAGGCGCGCGGTAATCATACGGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCACATGAGGCGCGCGGTAATCATACGGCGGCAATAGTTCCGGC 120
OY 121 CCGGACTCAACGTTGAGCATTTATCATGAGGTTCCGTAACGCTGCGCTTCTGCAA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCATGAGGTTCCGTAACGCTGCGCTTCTGCAA 180
```


Db 121 CCGGACTCAAGCTGAGATTATATGACGCTTCCGTTACGCTGGCTGCTCTGAA 180
Qy 181 AGCGATCCCGCTAAATCTGAAACGACATTAACGAGCGGTTATGTAACGGCCGAT 240
Db 181 AGCGATCCCGCTAAATCTGAAACGACATTAACGAGCGGTTATGTAACGGCCGAT 240
Qy 241 GTAGGCCGAGGCTGGCGATTAATGTAATCTGTAAGTCAAGATGCTTTCAGAAATAT 300
Db 241 TATATATCAGCTGTTATCCCGCTGTGTATCCATGAAATGGCATGATTCAGAAATAT 300
Qy 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Db 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Qy 361 AATAACCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTGTTTACCATGAAATG 420
Db 361 AATAACCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTGTTTACCATGAGTT 420
Qy 421 GCACATGCAACCAACGCGCAGGCTAACCATATTA 456
Db 421 GGTTTGGCAACCAACGCGCAGGCTAACCATATTA 456

RESULT 12

US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-29

Query Match 76.1%; Score 347.2; DB 23; Length 456;
Best Local Similarity 85.1%; Pred. No. 1e-100;
Matches 388; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACGCAATCGAGTTCTGGAGCTGCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACGCAATCGAGTTCTGGAGCTGCTGGCT 60
Qy 61 GGGCTGCTTCCACAATGGGGCGGCGGTATCATTAACGGCGCGGCAATAGTTCCGGC 120
Db 61 GGGCTGCTTCCACAATGGGGCGGCGGTATCATTAACGGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGCTGAGATTATATGACGCTTCCGTTACGCTGGCTGCTCTGAA 180
Db 121 CCGGACTCAAGCTGAGATTATATGACGCTTCCGTTACGCTGGCTGCTCTGAA 180
Qy 181 AGCGATCCCGCTAAATCTGAAACGACATTAACGAGCGGTTATGTAACGGCCGAT 240
Db 181 AGCGATCCCGCTAAATCTGAAACGACATTAACGAGCGGTTATGTAACGGCCGAT 240
Qy 241 GTAGGCCGAGGCTGGCGATTAATGTAATCTGTAAGTCAAGATGCTTTCAGAAATAT 300
Db 241 GTAGGCCGAGGCTGGCGATTAATGTAATCTGTAAGTCAAGATGCTTTCAGAAATAT 300

Qy 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Db 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Qy 361 AATAACCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTGTTTACCATGAAATG 420
Db 361 AATAACCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTGTTTACCATGAGTT 420
Qy 421 GCACATGCAACCAACGCGCAGGCTAACCATATTA 456
Db 421 GGTTTGGCAACCAACGCGCAGGCTAACCATATTA 456

RESULT 13

US-08-233-642A-54
Sequence 54, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-642A-54

Query Match 69.1%; Score 315.2; DB 6; Length 361;
Best Local Similarity 99.1%; Pred. No. 2e-90;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 64 GTGCTTCCACAATGGGGCGGCGGTAATCATTAACGGCGGCAATAGTTCCGGCCG 123
Db 1 GTGCTTCCACAATGGGGCGGCGGTAATCATTAACGGCGGCAATAGTTCCGGCCG 60
Qy 124 GACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCGCTTCTGCAAGC 183
Db 61 GACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCGCTTCTGCAAGC 120

```

QY 184 GATGCCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTAGCGCCGATGTA 243
|
|
|
Db 121 GATGCCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTAGCGCCGATGTA 180
|
|
|
QY 244 GGCAGCGGTGCGGATTAATGACTTATTTGAAGTCACTGAGATGTTTTCAGAAATATGCC 303
|
|
|
Db 181 GGCAGCGGTGCGGATTAATGACTTATTTGAAGTCACTGAGATGTTTTCAGAAATATGCC 240
|
|
|
QY 304 ACCATGACCAATGAGAACCTTAAATCTCCGATATTAATCTGTCGSCCAATACGGCGATAT 363
|
|
|
Db 241 ACCATGACCAATGAGAACCTTAAATCTCCGATATTAATCTGTCGSCCAATACGGCGATAT 300
|
|
|
QY 364 AACGCCGCGCTGTATTAATTA 383
|
|
|
Db 301 AACGCCGCGCTGTATTAATTA 320
|
|
|

```

RESULT 14
US-09-252-691-172
Sequence 172, Application US/09252691B

```

; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; PRIORITY FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-172

```

Query Match 48.8%; Score 222.4; DB 17; Length 477;
Best Local Similarity 68.0%; Pred. No. 1.9e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 60
|
|
|
Db 22 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 81
|
|
|
QY 61 GCGCTGTTCCACATGCGGCGCGGCGGTATCATTAACGCGCGCAATAGTTCCGCG 120
|
|
|
Db 82 GCGCTGTTCCACATGCGGCGCGGCGGTATCATTAACGCGCGCAATAGTTCCGCG 141
|
|
|
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTACGCTGCGCTTCTGCA 180
|
|
|
Db 142 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTACGCTGCGCTTCTGCA 201
|
|
|
QY 181 AGCGATGCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 240
|
|
|
Db 202 AGCGATGCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 261
|
|
|
QY 241 GTAGGCGAGGTTGCGGATTAATGATCTATTTGAATCTGACTGAGATGTTTCAAGAAATAT 300
|
|
|
Db 262 GTAGGCGAGGTTGCGGATTAATGATCTATTTGAATCTGACTGAGATGTTTCAAGAAATAT 321
|
|
|
QY 301 GCCACATGACCAAGTGGAAAGCTTAAATCTCCGATATTAATCTCCGCAATACGGCGGT 360
|
|
|
Db 322 GCCACATGACCAAGTGGAAAGCTTAAATCTCCGATATTAATCTCCGCAATACGGCGGT 381
|
|
|
QY 361 AATAAGCGCGGCTGTATTAATGATCTGAGTGTATACCGGTTGTTACCCATGAATG 420
|
|
|
Db 382 GGCAGCGCGCGGCGGTAGACGAGAGCGGTCCGCTCAACGTTGACCGAGTT 441
|
|
|
QY 421 GCACATGCAACCAACGCGCAAGGCTAACGATATTA 456
|
|
|
Db 442 GCGTTTGGCAACCAACGCGCAAGGCTAACGATATTA 477
|
|
|

```

RESULT 15
US-09-252-691C-172

```

; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-172

```

Query Match 48.8%; Score 222.4; DB 17; Length 477;
Best Local Similarity 68.0%; Pred. No. 1.9e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 60
|
|
|
Db 22 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 81
|
|
|
QY 61 GCGCTGTTCCACATGCGGCGCGGCGGTATCATTAACGCGCGCAATAGTTCCGCG 120
|
|
|
Db 82 GCGCTGTTCCACATGCGGCGCGGCGGTATCATTAACGCGCGCAATAGTTCCGCG 141
|
|
|
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTACGCTGCGCTTCTGCA 180
|
|
|
Db 142 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTACGCTGCGCTTCTGCA 201
|
|
|
QY 181 AGCGATGCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 240
|
|
|
Db 202 AGCGATGCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 261
|
|
|
QY 241 GTAGGCGAGGTTGCGGATTAATGATCTATTTGAATCTGACTGAGATGTTTCAAGAAATAT 300
|
|
|
Db 262 GTAGGCGAGGTTGCGGATTAATGATCTATTTGAATCTGACTGAGATGTTTCAAGAAATAT 321
|
|
|
QY 301 GCCACATGACCAAGTGGAAAGCTTAAATCTCCGATATTAATCTCCGCAATACGGCGGT 360
|
|
|
Db 322 GCCACATGACCAAGTGGAAAGCTTAAATCTCCGATATTAATCTCCGCAATACGGCGGT 381
|
|
|
QY 361 AATAAGCGCGGCTGTATTAATGATCTGAGTGTATACCGGTTGTTACCCATGAATG 420
|
|
|
Db 421 GCACATGCAACCAACGCGCAAGGCTAACGATATTA 456
|
|
|
Db 442 GCGTTTGGCAACCAACGCGCAAGGCTAACGATATTA 477
|
|
|

```

Search completed: March 18, 2004, 02:46:58
Job time : 2923.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)

2506.048 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456

Sequence: 1 atgaactttaaagtgc.....ccacgctaaccagtataa 456

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA New:*
1: /cgn2_6/ptodata/2/pna/PCT NEW COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06 NEW COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07 NEW COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08 NEW COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09 NEW COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10 NEW COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60 NEW COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	216	47.4	675	5 US-09-741-873C-3	Sequence 3, Appl1
2	172	37.7	396	5 US-09-741-873C-1	Sequence 1, Appl1
3	33.4	7.3	3990	5 US-09-830-230A-629	Sequence 629, App
4	33.2	7.3	1355	6 US-10-045-674A-593	Sequence 593, App
5	32.4	7.1	150223	6 US-10-624-149A-1	Sequence 1, Appl1
6	31.4	6.9	1588	6 US-10-767-701-14522	Sequence 14522, A
7	31.2	6.8	280892	6 US-10-767-471-10809	Sequence 10809, A
8	30.2	6.6	512	6 US-10-781-014-257	Sequence 257, App
9	30.2	6.6	1326	6 US-10-781-014-255	Sequence 255, App
10	29.6	6.5	24900	6 US-10-767-471-10736	Sequence 10736, A
11	29.4	6.4	300	6 US-10-767-701-28437	Sequence 28437, A
12	29.4	6.4	1400	6 US-10-545-213-2134	Sequence 2134, App
13	29.4	6.4	1400	7 US-60-545-213-2135	Sequence 2135, App
14	29.4	6.4	1400	7 US-60-545-213-6406	Sequence 6406, App
15	29.4	6.4	1400	7 US-60-545-213-6407	Sequence 6407, App
16	28.6	6.3	1186	1 PCT-US04-05654-1030	Sequence 1030, App
17	28.4	6.2	1350	6 US-10-784-480-291	Sequence 291, App
18	28.4	6.2	394468	6 US-60-548-091-5725	Sequence 5725, App
19	28.2	6.2	596	6 US-10-767-701-5378	Sequence 5378, App
20	28.2	6.2	610	6 US-10-767-701-8781	Sequence 8781, App
21	28.2	6.2	1746	1 PCT-US04-05654-2555	Sequence 2555, App
22	28	6.1	237	6 US-10-767-701-31417	Sequence 31417, A
23	28	6.1	600	7 US-60-545-213-1313	Sequence 1313, App
24	28	6.1	600	7 US-60-545-213-1314	Sequence 1314, App
25	28	6.1	600	7 US-60-545-213-5585	Sequence 5585, App
26	28	6.1	600	7 US-60-545-213-5586	Sequence 5586, App

ALIGNMENTS

C	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
C	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
C	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
C	30	27.8	6.1	1580	6	US-10-451-467A-239	Sequence 239, App
C	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
C	32	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
C	33	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, App
C	34	27.6	6.1	6224	6	US-10-453-372-1187	Sequence 1187, App
C	35	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, App
C	36	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
C	37	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, App
C	38	27.4	6.0	1295	6	US-10-100-683-5480	Sequence 5480, App
C	39	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appl1
C	40	27.4	6.0	3416	6	US-10-101-698A-3397	Sequence 3397, App
C	41	27.4	6.0	8206	6	US-10-021-698A-3529	Sequence 3529, App
C	42	27.4	6.0	186449	6	US-10-021-698A-715	Sequence 715, App
C	43	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, App
C	44	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, App
C	45	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appl1

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C

GENERAL INFORMATION:
APPLICANT: Normark, Staffan

TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

Db 203 CCAATTCTGACGTGAACATTTTACCAATGACGTGGCGGTAACTGTGACTTGTCTTGCA 262
QY 181 AGCGATGCCCGGTAATCTGAAACGACATTAACCAAGCGGTTATGTTAAGCGCGCAT 240
Db 263 ACTGATGCCCGGTAATCTGACCTTACTATTTACCAAGATGCGCGGTAATGTTGACAT 240
QY 241 GTAGCCAGCGGTGGGTAATATGATCTATTTGAACCTGACGATGTTGTTGCAAAATAT 322
Db 323 GTTGGTCAAGGCTGACATGACGCTCAATGATCTGACCAAGTGGTTCGGTAAACGC 300
QY 301 GCCACATCGACCGATGGAACGCTAAATACTCCGATATTAATCTGCGGCAATACGCGGT 382
Db 383 GCTACTCTTATGATGCTGGAACCGCAAAATTTGGAATGACGTTAAACAGTTGCGTGT 360
QY 361 AATAACGCCCGGCTGGTTATTAATGATCACTGCTGTTACCGGTGTTACCATGAATG 442
Db 443 GGCACCGGTGCTGCAATTTGACCACTGATCTAACTCTCGTCAAGTACATCAGATT 420
QY 421 GCACATGCAACACGCGCAACGCTAACCATGATTA 456
Db 503 GCGCTTGTGTAACACGCGACCGCTCATGACTACTA 538

RESULT 2
US-09-741-873C-1
Sequence 1, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22-1
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06-878
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 396
TYPE: DNA
ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 37.7%; Score 172; DB 5; Length 396;
Best Local Similarity 64.6%; Pred. No. 1.7e-47;
Matches 256; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 61 GGCCTGCTTCAACAATGGGCGCGGGGATCATATACGCGCGGCAATATGTTCCGGC 120
Db 1 GGTGTGTCTCTGATGAGCGCGCGGATTAACACGCGGTGGGATTAATATAGCGGC 60
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGGTTCGGTAACGCTGCGCTGCTTCCAA 180
Db 61 CCAATTTCTAGCTGAACATTATCCAGTACGTTGCGGATCTGCACTTGTCTGCAA 120
QY 181 AGCGATGCCCGGTAATCTGAAACGACATTAACCAAGCGGTTATGTTAAGCGCGCAT 240
Db 121 ACTGATGCCCGGTAATCTGACCTTACTATTTACCAAGATGCGCGGTAATGTTGACAT 180
QY 241 GTAGCCAGCGGTGGGTAATATGATCTATTTGAACCTGACGATGTTTCAAAATAT 300

Db 181 GTTGTGACGGCTCAGATGACAGCTCAATGCATGACCAAGCGCTTGGTAAACGC 240
QY 301 GCCACATCGACCGATGGAACGCTAAATACTCCGATATTAATCTGCGGCAATACGCGGT 360
Db 241 GCTACTCTTATGATGCTGGAACCGCAAAATTTGGAATGACGTTAAACAGTTGCGTGT 300
QY 361 AATAACGCCCGGCTGGTTATTAATGATCACTGCTGTTACCGGTGTTACCATGAATG 420
Db 301 GGCACCGGTGCTGCAATTTGACCACTGATCTAACTCTCGTCAAGTACATCAGATT 360
QY 421 GCACATGCAACACGCGCAACGCTAACCATGATTA 456
Db 361 GCGCTTGTGTAACACGCGACCGCTCATGACTACTA 396

RESULT 3
US-09-830-230A-629
Sequence 629, Application US/09830230A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Lyme Disease Vaccines
FILE REFERENCE: PB481US
CURRENT APPLICATION NUMBER: US/09/830,230A
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: PCT/US98/12718
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/057,483
PRIOR FILING DATE: 1997-09-03
PRIOR APPLICATION NUMBER: 60/053,344
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/053,377
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/050,359
PRIOR FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 756
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 629
LENGTH: 3990
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1135)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1143)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1148)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1210)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1245)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1247)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1250)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1251)
OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

```

RESULT 4
US-10-045-674A-593
: Sequence 593, Application US/10045674A
: GENERAL INFORMATION:
: APPLICANT: LADNER, ROBERT C.
: APPLICANT: COHEN, EDWARD H.
: APPLICANT: NASTRI, HORACIO G.
: APPLICANT: ROOKEY, KRISTIN L.
: APPLICANT: HOET, KENE
: APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
: TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
: TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
: TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
: TITLE OR INVENTION: LIBRARIES
: FILE REFERENCE: DYAX/002 CIP2
: CURRENT APPLICATION NUMBER: US/10/045,674A
: CURRENT FILING DATE: 2001-10-25
: PRIOR APPLICATION NUMBER: 06/198,069
: PRIOR FILING DATE: 2000-04-17
: PRIOR APPLICATION NUMBER: 09/837,306
: PRIOR FILING DATE: 2001-04-17
: NUMBER OF SEQ ID NOS: 635
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 593
: LENGTH: 1355
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: M13-III
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1305)
: US-10-045-674A-593

```

OY 258 TAATAGTCTATTGTAACGTACGCACAAGAAGTTCCGAATAATATGCCACCATGACCAGTG 317
Dd 948 RYTTCGAAWVSVCIGCYACWGANATAYAGTGCTGCVCATCGAIVGGYTTYATATYGAGVATPYS 100
OY 318 GAACGCTAAAACACTCCGATATTATCTGTGCGGCCCAATACGGCGGGTAAATPAACGCGCGTGGA 377
Dd 1008 YGGICTKCTCTAAYGGIYAAYGGMGCTACYGCGMAYYTTTGGCWGSITCKAATATTCARATGGC 106
OY 378 TAAATTATGATCAGCTGTTACCCGCTGTGTTAACCATGAAATGGGACA 425
Dd 1068 YCAGTYGGMGAAYGGKGAIYAAYMSCKRTTAATGAIAAYATTMYMCA 1115

```

Query Match      7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 3.5;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY      173 CTCTGCAAAAGCGATGCCCGCTAAATCTGAACGACCTAACCCAGACGGTTATGTGTAACG 232
Db      95542 CTGTGGTAACGATGGGGCGCTTAGTTTGAATACGGCTCATTTCCACAGCAGATTTGGCCTAACG 95601

QY      233 GCGCGCGATGTAGGSCCAAGGGGCGCGATTAATAGTACTATTGGACTGACTCGAATAGTTTCA 292
Db      95602 TCAAGCCCATGAGCCCAAGCAGACCGGATATGTGTTCTGTATACAGTCTCTGGGGCGCGTCA 95661

QY      293 GAAATTAATGCCACCATCGACCACTGG 318
Db      95662 GAGAGCTTATTCAACAAAGAGCTGTGG 95687

RESULT 6
US-10-767-701-14522
; Sequence 14522; Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

```

[illegible]

Query Match	6.6%	Score 30.2	DB 6	Length 512
Best Local Similarity	56.6%	Pred. No. 1.8		
Matches	56	Conservative	0	Mismatches 43
				Indels 0
				Gaps 0
QY	310	GACACGTGGAAGCGTAAATACTCGATATATCTGTGGCCATAACGGCGGTATAACGCC	369	
Db	375	GCCCATGTCACCGCCGACACACCACTTATATATATATATGCGCTCTTCAGCAGCCGAAAGGCG	316	
QY	370	GCGCTGGTTAATATATATGATCAAGCTGTTACCGGTGTGTT	408	
Db	315	GTACTGCTTCAGCATGTCCTCGTGTGTGAGAGGTGAT	277	

RESULT 9
US-10-761-014-255/c
Sequence 255, Application US/10761014
GENERAL INFORMATION:
APPLICANT: Pompejus, Markus
APPLICANT: Krogger, Burkhard
APPLICANT: Schroder, Hartwig
APPLICANT: Zelder, Oskar
APPLICANT: Haberer, Gregor
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY
FILE REFERENCE: BGI-126PCN
CURRENT APPLICATION NUMBER: US/10/761,014
CURRENT FILING DATE: 2004-02-17
PRIOR APPLICATION NUMBER: US 09/602,740
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/141,031
PRIOR FILING DATE: 1999-06-25
PRIOR APPLICATION NUMBER: 60/143,208
PRIOR FILING DATE: 1999-07-09
PRIOR APPLICATION NUMBER: 60/151,572
PRIOR FILING DATE: 1999-08-31
PRIOR APPLICATION NUMBER: DE 19931412.8
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931413.6
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931419.5
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931420.9
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931424.1
PRIOR FILING DATE: 1999-07-08
PRIOR APPLICATION NUMBER: DE 19931428.4
PRIOR FILING DATE: 1999-07-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 784
SEQ ID NO 255
LENGTH: 1326
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
FEATURE:
NAME/KEY: CDS
LOCATION: (101)..(1303)
OTHER INFORMATION: RXN02596
US-10-761-014-255

Query Match 6.6%; Score 30.2; DB 6; Length 1326;
Best Local Similarity 56.6%; Pred. No. 2.7;
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 310 GACGAGTGAACGCTAAACCTCCGATTAATCTGCGGCAATACGGCGTAAATACGCC 369
Db 1189 GCCAGGTGACCGCGGACACACACTTATTAAGACACCTCTTCAGACGACGAGCG 1130

Qy 370 GCGGTGTTATTAATGATCAGCTGTTACCCGTTGTT 408
Db 1129 GTAAGCTTTCAGCATGCTCGGTGAGTGTGAT 1091

RESULT 10
US-10-767-471-10736/c
Sequence 10736, Application US/10767471
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01505
CURRENT APPLICATION NUMBER: US/10/767,471
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10736
LENGTH: 24900
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(24900)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

Query Match 6.5%; Score 29.6; DB 6; Length 24900;
Best Local Similarity 52.4%; Pred. No. 14;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 207 CATTACCCAGAGCGGTTATGTTACGCGCCGATGAGCCAGGTCGGATTAATGATAC 266
Db 10048 CAGCATCTGACTGCTTTAGTGTCTCAAAAGTCTGCTCACTGTCATTAATGCTG 9989

Qy 267 TATTGAATGCTGAGATGTTTCAGAAATTAATGCCATGACGACGAGTGAACGCTAA 326
Db 9988 TCCTGAACCCAACTGATGTAATTAATTAATGATGCAATCTGTTGTTATATTTAA 9929

Qy 327 AAC 330
Db 9928 AATC 9925

RESULT 11
US-10-767-701-28437
Sequence 28437, Application US/10767701
GENERAL INFORMATION:
APPLICANT: KOVALLIC, David K.
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(5335)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 28437
LENGTH: 300
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437

Query Match 6.4%; Score 29.4; DB 6; Length 300;
Best Local Similarity 54.1%; Pred. No. 2.7;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 14 AAGTGACGAGTTCGACGACATCGTATGTTCTGACAGTGTCTGCGCTGCTCCAC 73
Db 123 AAGAAAGCGGTTACGAAGCGGTGTGATGAGTGGCGGCGGCGTTACGCGGCTCCC 182

Qy 74 AATGGCGCGCGCGGTTAATCATACGCGCGGCAATAGTCCGCGCGG 124
Db 183 GTAAGCGCGCGCGGTTATGCGCGCGCGCGGTTACGCGTACGCAACCGG 233

RESULT 12
US-60-545-213-2134/c
Sequence 2134, Application US/60545213
GENERAL INFORMATION:
APPLICANT: Wyeth
APPLICANT: Mounts, William Martin
TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
FILE REFERENCE: AM101083 (031896-042099)
CURRENT APPLICATION NUMBER: US/60/545,213
CURRENT FILING DATE: 2004-02-18
NUMBER OF SEQ ID NOS: 303284

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2134

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-2134

Query Match

Best Local Similarity 45.2%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTGTAACGCGCCGATGTAGCCAGGTTGGCATATATGACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGCAGGTTGGGTGATCGGTTATTA 378

272 AACTGACTCAGATGTCTTCAGAAATTAATGCCACCATGACCAAGTGAACGCTAAACT 331

377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGCAATCTCTCACTCTGAACT 318

332 CCGATATTACTGTGCGCCCAATACGGCGGTATTAACCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGACGAGCGGCTCTGATGCAATTAACCTGCTGATCTGCTCCTCAGGTTTCT 258

392 TGGTTACCGGTGTGTTATCCCATGAATGCGACATGCAACAGCCACGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTGTGTTATTTCTGAAGTGAATGAAGAGGGGTGTGACTG 199

RESULT 13

US-60-545-213-2135/c

Sequence 2135, Application US/60545213

GENERAL INFORMATION:

APPLICANT: Wyeth

APPLICANT: Mounts, William Martin

TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug

FILE REFERENCE: AM101083 (031896-042099)

CURRENT APPLICATION NUMBER: US/60/545,213

CURRENT FILING DATE: 2004-02-18

NUMBER OF SEQ ID NOS: 303284

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2135

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-2135

Query Match

Best Local Similarity 45.2%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTGTAACGCGCCGATGTAGCCAGGTTGGCATATATGACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGCAGGTTGGGTGATCGGTTATTA 378

272 AACTGACTCAGATGTCTTCAGAAATTAATGCCACCATGACCAAGTGAACGCTAAACT 331

377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGCAATCTCTCAGCTCTGAACT 318

332 CCGATATTACTGTGCGCCCAATACGGCGGTATTAACCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGACGAGCGGCTCTGATGCAATTAACCTGCTGATCTGCTCCTCAGGTTTCT 258

392 TGGTTACCGGTGTGTTATCCCATGAATGCGACATGCAACAGCCACGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTGTGTTATTTCTGAAGTGAATGAAGAGGGGTGTGACTG 199

RESULT 14

US-60-545-213-6406/c

Sequence 6406, Application US/60545213

GENERAL INFORMATION:

APPLICANT: Wyeth

APPLICANT: Mounts, William Martin

TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug

FILE REFERENCE: AM101083 (031896-042099)

CURRENT APPLICATION NUMBER: US/60/545,213

CURRENT FILING DATE: 2004-02-18

NUMBER OF SEQ ID NOS: 303284

SOFTWARE: PatentIn version 3.2

SEQ ID NO 6406

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-6406

Query Match

Best Local Similarity 45.2%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTGTAACGCGCCGATGTAGCCAGGTTGGCATATATGACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGCAGGTTGGGTGATCGGTTATTA 378

272 AACTGACTCAGATGTCTTCAGAAATTAATGCCACCATGACCAAGTGAACGCTAAACT 331

377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGCAATCTCTCAGCTCTGAACT 318

332 CCGATATTACTGTGCGCCCAATACGGCGGTATTAACCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGACGAGCGGCTCTGATGCAATTAACCTGCTGATCTGCTCCTCAGGTTTCT 258

392 TGGTTACCGGTGTGTTATCCCATGAATGCGACATGCAACAGCCACGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTGTGTTATTTCTGAAGTGAATGAAGAGGGGTGTGACTG 199

RESULT 15

US-60-545-213-6407/c

Sequence 6407, Application US/60545213

GENERAL INFORMATION:

APPLICANT: Wyeth

APPLICANT: Mounts, William Martin

TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug

FILE REFERENCE: AM101083 (031896-042099)

CURRENT APPLICATION NUMBER: US/60/545,213

CURRENT FILING DATE: 2004-02-18

NUMBER OF SEQ ID NOS: 303284

SOFTWARE: PatentIn version 3.2

SEQ ID NO 6407

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-6407

Query Match

Best Local Similarity 45.2%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTGTAACGCGCCGATGTAGCCAGGTTGGCATATATGACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGCAGGTTGGGTGATCGGTTATTA 378

272 AACTGACTCAGATGTCTTCAGAAATTAATGCCACCATGACCAAGTGAACGCTAAACT 331

377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGCAATCTCTCAGCTCTGAACT 318

332 CCGATATTACTGTGCGCCCAATACGGCGGTATTAACCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGACGAGCGGCTCTGATGCAATTAACCTGCTGATCTGCTCCTCAGGTTTCT 258

392 TGGTTACCGGTGTGTTATCCCATGAATGCGACATGCAACAGCCACGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTGTGTTATTTCTGAAGTGAATGAAGAGGGGTGTGACTG 199

Thu Mar 18 12:27:56 2004

us-09-543-407-11.rnpn

Page 7

Search completed: March 17, 2004, 08:25:40
Job time : 71.6301 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 : Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456
Sequence: 1 atgaacttttaaaagtgc.....ccacgcttaaccagtataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA Main:*

- 1: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 8: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 9: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 10: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 11: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 12: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 13: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 14: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 15: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 16: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 17: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 18: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 19: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 20: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 21: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 22: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 23: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 24: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 25: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 26: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 27: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 28: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 29: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 30: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 31: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 32: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 33: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 34: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 35: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 36: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 37: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 38: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 39: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 40: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 41: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 42: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
- 43: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*

- 44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 45: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 46: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 47: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 48: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 49: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 50: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 51: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 52: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 53: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 54: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 55: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 56: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 57: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 58: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 59: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 60: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 61: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 62: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 63: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 64: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 65: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 66: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 67: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 68: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 69: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 70: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 71: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 72: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 73: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 74: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 75: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 76: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 77: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 78: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 79: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 80: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 81: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 82: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 83: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 84: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 85: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 86: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 87: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 88: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 89: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 90: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 91: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 92: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 93: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 94: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 95: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 96: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 97: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 98: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 99: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 100: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 101: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 102: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 103: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 104: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 105: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 106: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 107: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 108: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 109: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 110: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-13	Sequence 13, Appl
2	404.8	88.8	456	6	US-08-233-642A-56	Sequence 56, Appl
3	404.8	88.8	456	23	US-09-543-407-1	Sequence 1, Appl1
4	382.4	83.9	456	23	US-09-543-407-23	Sequence 23, Appl
5	376	82.5	456	23	US-09-543-407-11	Sequence 11, Appl
6	361.6	79.3	456	23	US-09-543-407-19	Sequence 19, Appl
7	355.2	77.9	456	23	US-09-543-407-17	Sequence 17, Appl
8	353.6	77.5	456	23	US-09-543-407-25	Sequence 25, Appl
9	352	77.2	456	23	US-09-543-407-15	Sequence 15, Appl
10	350.4	76.8	456	23	US-09-543-407-21	Sequence 21, Appl
11	350.4	76.8	456	23	US-09-543-407-27	Sequence 27, Appl
12	348.8	76.5	456	23	US-09-543-407-29	Sequence 29, Appl
13	287.2	63.0	361	6	US-08-233-642A-54	Sequence 54, Appl
14	227.2	49.8	477	17	US-09-252-691C-172	Sequence 172, Appl
15	227.2	49.8	477	17	US-09-252-691C-172	Sequence 172, Appl
16	227.2	49.8	477	51	US-10-417-886-172	Sequence 172, Appl
17	225.6	49.5	456	23	US-09-543-407-3	Sequence 3, Appl1
18	225.6	49.5	675	13	US-08-978-878-3	Sequence 3, Appl1
19	225.6	49.5	675	13	US-09-741-873B-3	Sequence 3, Appl1
20	223.2	48.9	453	92	US-60-352-946-1	Sequence 1, Appl1
21	223.2	48.9	453	101	US-60-444-371-1	Sequence 1, Appl1
22	181.6	39.8	396	13	US-08-978-878-1	Sequence 1, Appl1
23	156.8	34.4	360	31	US-09-741-873B-1	Sequence 1, Appl1
24	49	10.7	78	23	US-09-543-407-53	Sequence 53, Appl
25	48.4	10.6	78	23	US-09-543-407-52	Sequence 52, Appl
26	48	10.5	48	23	US-09-543-407-9	Sequence 9, Appl1
27	44.8	9.8	2000	36	US-09-887-372A-5263	Sequence 5263, Ap
28	41.6	9.1	78	23	US-09-543-407-48	Sequence 48, Appl
29	40.8	8.9	100	45	US-10-146-492B-79	Sequence 79, Appl
30	40.6	8.9	78	23	US-09-543-407-49	Sequence 49, Appl
31	39.4	8.6	2732	51	US-10-437-963-37426	Sequence 37426, A
32	37.6	8.2	456	23	US-09-543-407-2	Sequence 2, Appl1
33	37.6	8.2	522	17	US-09-252-691C-170	Sequence 170, Appl
34	37.6	8.2	522	17	US-09-252-691C-170	Sequence 170, Appl
35	37.6	8.2	522	17	US-10-417-886-170	Sequence 170, Appl
36	37.4	8.2	456	23	US-09-543-407-4	Sequence 4, Appl1
37	37.4	8.2	1304	45	US-10-437-963-19773	Sequence 19773, A
38	35.4	7.8	100	45	US-10-146-492B-77	Sequence 77, Appl
39	35.4	7.8	3411	1	PCT-US01-08631-24330	Sequence 24330, A
40	35.4	7.7	100	45	US-10-146-492B-78	Sequence 78, Appl
41	34.8	7.6	1083	1	PCT-US01-08631-12549	Sequence 12549, A
42	34.8	7.5	2751	50	US-10-398-221-3563	Sequence 3563, Ap
43	34.2	7.5	385	18	US-09-333-535A-288	Sequence 288, App
44	34.2	7.5	385	25	US-09-615-606A-909	Sequence 909, App

ALIGNMENTS

RESULT 1
 US-09-543-407-13
 GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 APPLICANT: Kay, William W.
 TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 FILE REFERENCE: 920043.406
 CURRENT APPLICATION NUMBER: US/09/543.407
 CURRENT FILING DATE: 2000-04-05
 NUMBER OF SEQ. ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ. ID NO 13
 LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence

FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b atga
 OTHER INFORMATION: sequence containing the replacement fragment
 OTHER INFORMATION: encoding PT3 from GPe3 of *Leishmania* major.
 US-09-543-407-13

Query Match 100.0%; Score 456; DB 23; Length 456;
 Best Local Similarity 100.0%; Pred. No. 7.1e-135;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATAACTTTAAATGAGATTCGACGAATGTTGTTGCGAGTGTCTGGCT 60
   |||
DB 1 ATAACTTTAAATGAGATTCGACGAATGTTGTTGCGAGTGTCTGGCT 60
   |||
QY 61 GCGGTCTTCAATGAGGCGCGCGGTTAATCATACGCGCGCATATGTTCCGC 120
   |||
DB 61 GCGGTCTTCAATGAGGCGCGCGGTTAATCATACGCGCGCATATGTTCCGC 120
   |||
QY 121 CCGGACTCAAGTTGAGATTATTCAGTACGTTCCGTTACGCTGCTGCAA 180
   |||
DB 121 CCGGACTCAAGTTGAGATTATTCAGTACGTTCCGTTACGCTGCTGCAA 180
   |||
QY 181 AGGATGCGCGTAATCTGAACGACCATACCGAGCGTTATGTAACGCGCGAT 240
   |||
DB 181 AGGATGCGCGTAATCTGAACGACCATACCGAGCGTTATGTAACGCGCGAT 240
   |||
QY 241 GTAGGCGAGGTCGCGATATGTAATGTAATGTAATGTAATGTAATGTAAT 300
   |||
DB 241 GTAGGCGAGGTCGCGATATGTAATGTAATGTAATGTAATGTAATGTAAT 300
   |||
QY 301 GCGACCATGACGATGAGGACGTAATGTAATGTAATGTAATGTAATGTAAT 360
   |||
DB 301 GCGACCATGACGATGAGGACGTAATGTAATGTAATGTAATGTAATGTAAT 360
   |||
QY 361 CCGGTTACCGGTTGTTATCCATGAAATGGAACGATGATGTTGTTGTTGTT 420
   |||
DB 361 CCGGTTACCGGTTGTTATCCATGAAATGGAACGATGATGTTGTTGTTGTT 420
   |||
QY 421 GGTGTTGCAACGACGCGCTAATGTAATGTAATGTAATGTAATGTAAT 456
   |||
DB 421 GGTGTTGCAACGACGCGCTAATGTAATGTAATGTAATGTAATGTAAT 456
   |||

```

RESULT 2
 US-08-233-642A-56
 GENERAL INFORMATION:
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, S. Karen
 APPLICANT: Clouthier, Sharon C.
 APPLICANT: Doran, James L.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
 BASED VACCINES
 NUMBER OF SEQUENCES: 58
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233.642A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 424
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 88.8%; Score 404.8; DB 6; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.9e-118;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT 60
DB 1 ATGAACTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT 60
QY 61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
DB 61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTGCTCTGCAA 180
QY 181 AGCGATGCCGTAAATCTGAAACGACCATTAACCAAGCGGTATAGTAAACGCGCGAT 240
DB 181 AGCGATGCCGTAAATCTGAAACGACCATTAACCAAGCGGTATAGTAAACGCGCGAT 240
QY 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCGTCAAGATGTTTGAAGAAATAT 300
DB 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCGTCAAGATGTTTGAAGAAATAT 300
QY 301 GCCACATTCGACGAGTGAACGCTAAAACTCCGATTTACTGTCGGCAATATGATCAG 360
DB 301 GCCACATTCGACGAGTGAACGCTAAAACTCCGATTTACTGTCGGCAATATGATCAG 360
QY 361 CTGGTACCGGTGTGTTTATCCCATGAAATGGCAATGCAACGCTATGCTGCTGAGTT 420
DB 361 AATAACCGCGCGTGTATATCAGACCGCATCTGATTCACAGTAATGTCGTCAGGTT 420
QY 421 GCTTTGGCAACAACGCGCAAGGCTAACCAATATTA 456
DB 421 GCTTTGGCAACAACGCGCAAGGCTAACCAATATTA 456

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1

Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.9e-118;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT 60
DB 1 ATGAACTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT 60
QY 61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
DB 61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTGCTCTGCAA 180
QY 181 AGCGATGCCGTAAATCTGAAACGACCATTAACCAAGCGGTATAGTAAACGCGCGAT 240
DB 181 AGCGATGCCGTAAATCTGAAACGACCATTAACCAAGCGGTATAGTAAACGCGCGAT 240
QY 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCGTCAAGATGTTTGAAGAAATAT 300
DB 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCGTCAAGATGTTTGAAGAAATAT 300
QY 301 GCCACATTCGACGAGTGAACGCTAAAACTCCGATTTACTGTCGGCAATATGATCAG 360
DB 301 GCCACATTCGACGAGTGAACGCTAAAACTCCGATTTACTGTCGGCAATATGATCAG 360
QY 361 CTGGTACCGGTGTGTTTATCCCATGAAATGGCAATGCAACGCTATGCTGCTGAGTT 420
DB 361 AATAACCGCGCGTGTATATCAGACCGCATCTGATTCACAGTAATGTCGTCAGGTT 420
QY 421 GCTTTGGCAACAACGCGCAAGGCTAACCAATATTA 456
DB 421 GCTTTGGCAACAACGCGCAAGGCTAACCAATATTA 456

RESULT 4
US-09-543-407-23
Sequence 23, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding p13 from Gp63 of Leishmania major.

US-09-543-407-23

Query Match 83.9%; Score 382.4; DB 23; Length 456;
Best Local Similarity 89.9%; Pred. No. 2.8e-111;
Matches 410; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT 60
DB 1 ATGAACTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT 60
QY 61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120

Db 61 GGGCGTTCACAAATGGGCGCGCGCTAATACAGCGCGCAATAGTCCGGC 120
Qy 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCCGCTAACGCTGGCTTGTCTGCAA 180
Db 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCCGCTAACGCTGGCTTGTCTGCAA 180
Qy 181 AGCGATCCCGGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Db 181 AGCGATCCCGGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Qy 241 GTAGCCAGGCGTGGATTAATAGTACTATTTGAATGAACTGAAATGTTTCAAGAAATAT 300
Db 241 GTAGCCAGGCGTGGATTAATAGTACTATTTGAATGAACTGAAATGTTTCAAGAAATAT 300
Qy 301 GCCACATGACAGCAGTGAAGCGTAAACCTCCGATATTACTGTCCGCAATATGATCAG 360
Db 301 GCCACATGACAGCAGTGAAGCGTAAACCTCCGATATTACTGTCCGCAATATGATCAG 360
Qy 361 CTGCTTACCCGCTGTGTTACCCATGAATGCGACATGCAAGGTAATGCTGCTCAGGTT 420
Db 361 CATGAATGCGACATCAATCAATGACCGCATCTGATTCAGCGTAAATGCTGCTCAGGTT 420
Qy 421 GGTTTGGCAACACCGCGCTAACCGATTTAA 456
Db 421 GGTTTGGCAACACCGCGCTAACCGATTTAA 456

RESULT 5

US-09-543-407-11
Sequence 11, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043, 406
CURRENT APPLICATION NUMBER: US/09/543, 407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-11

Query Match 82.5%; Score 376; DB 23; Length 456;
Best Local Similarity 89.0%; Pred. No. 3,2e-109;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 1 ATGAACTTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Qy 61 GGGCTGCTTCCACAAATGGGCGCGCGGTAAATCATTAAGCGCGCAATAGTTCCGCG 120
Db 61 GGGCTGCTTCCACAAATGGGCGCGCGGTAAATCATTAAGCGCGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCCGCTAACGCTGGCTTGTCTGCAA 180
Db 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCCGCTAACGCTGGCTTGTCTGCAA 180
Qy 181 AGCGATCCCGGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Db 181 AGCGATCCCGGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240

RESULT 6

US-09-543-407-19
Sequence 19, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043, 406
CURRENT APPLICATION NUMBER: US/09/543, 407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-19

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 1.3e-104;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
Qy 1 ATGAACTTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Qy 61 GGGCTGCTTCCACAAATGGGCGCGCGGTAAATCATTAAGCGCGCAATAGTTCCGCG 120
Db 61 GGGCTGCTTCCACAAATGGGCGCGCGGTAAATCATTAAGCGCGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCCGCTAACGCTGGCTTGTCTGCAA 180
Db 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCCGCTAACGCTGGCTTGTCTGCAA 180
Qy 181 AGCGATCCCGGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Db 181 AGCGATCCCGGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Qy 241 GTAGCCAGGCGTGGATTAATAGTACTATTTGAATGAACTGAAATGTTTCAAGAAATAT 300
Db 241 GTAGCCAGGCGTGGATTAATAGTACTATTTGAATGAACTGAAATGTTTCAAGAAATAT 300
Qy 301 GCCACATGACAGCAGTGAAGCGTAAACCTCCGATATTACTGTCCGCAATATGATCAG 360
Db 301 GCCACATGACAGCAGTGAAGCGTAAACCTCCGATATTACTGTCCGCAATATGATCAG 360
Qy 361 CTGCTTACCCGCTGTGTTACCCATGAATGCGACATGCAAGGTAATGCTGCTCAGGTT 420
Db 361 CTGCTTACCCGCTGTGTTACCCATGAATGCGACATGCAAGGTAATGCTGCTCAGGTT 420

Qy 421 GGTITGGCAACAGCCAGCGCTAACAGATTAA 456
Db 421 GGTITGGCAACAGCCAGCGCTAACAGATTAA 456

RESULT 7

US-09-543-407-17
Sequence 17, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.9%; Score 355.2; DB 23; Length 456;

Best Local Similarity 86.2%; Pred. No. 1.5e-102; Indels 0; Gaps 0;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 ATGAACCTTTTAAAGTGGAGCATTCCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACCTTTTAAAGTGGAGCATTCCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTGTTCCACATGGGGGCGCGGCGGTATATCAATACGCGCGGCAATAGTTCCGGC 120
Db 61 GGCCTGTTCCACATGGGGGCGCGGCGGTATATCAATACGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGATTATCAGTACGTTCCGCTACGCTCGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGATTATCAGTACGTTCCGCTACGCTCGCTGCTGCA 180
Qy 181 AGCGATCCCGTAAATCTGAAGACCACTTACCGAGCGGTATGTAACGCGCGCAT 240
Db 181 AGCGATCCCGTAAATCTGAAGACCACTTACCGAGCGGTATGTAACGCGCGCAT 240
Qy 241 GTAGCCAGGGTGGGATTAATAGTACTTAAGTGAAGTGGTTCAGAAATAT 300
Db 241 GTAGCCAGGGTGGGATTAATAGTACTTAAGTGAAGTGGTTCAGAAATAT 300
Qy 301 GCCACATCGACAGTGAAGCTAAATCCCATATTAATGTCGCGCAATATGATCAG 360
Db 301 GCCACATCGACAGTGAAGCTAAATCCCATATTAATGTCGCGCAATATGATCAG 360
Qy 361 CTGGTACCCGTTGTTTACCATGAATGGCACATGCAACGTAATGTCGTCAGGTT 420
Db 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCACAGTAATGTCGTCAGGTT 420
Qy 421 GGTITGGCAACAGCCAGCGCTAACAGATTAA 456
Db 421 GGTITGGCAACAGCCAGCGCTAACAGATTAA 456

RESULT 8

US-09-543-407-25
Sequence 25, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

Query Match 77.5%; Score 353.6; DB 23; Length 456;

Best Local Similarity 86.0%; Pred. No. 4.8e-102; Indels 0; Gaps 0;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1 ATGAACCTTTTAAAGTGGAGCATTCCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACCTTTTAAAGTGGAGCATTCCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTGTTCCACATGGGGGCGCGGCGGTATATCAATACGCGCGGCAATAGTTCCGGC 120
Db 61 GGCCTGTTCCACATGGGGGCGCGGCGGTATATCAATACGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGATTATCAGTACGTTCCGCTACGCTCGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGATTATCAGTACGTTCCGCTACGCTCGCTGCTGCA 180
Qy 181 AGCGATCCCGTAAATCTGAAGACCACTTACCGAGCGGTATGTAACGCGCGCAT 240
Db 181 CTGGTACCCGTTGTTTACCATGAATGGCACATGCAACGTAATGTCGTCAGGTT 420
Qy 241 GTAGCCAGGGTGGGATTAATAGTACTTAAGTGAAGTGGTTCAGAAATAT 300
Db 241 GTAGCCAGGGTGGGATTAATAGTACTTAAGTGAAGTGGTTCAGAAATAT 300
Qy 301 GCCACATCGACAGTGAAGCTAAATCCCATATTAATGTCGCGCAATATGATCAG 360
Db 301 GCCACATCGACAGTGAAGCTAAATCCCATATTAATGTCGCGCAATATGATCAG 360
Qy 361 CTGGTACCCGTTGTTTACCATGAATGGCACATGCAACGTAATGTCGTCAGGTT 420
Db 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCACAGTAATGTCGTCAGGTT 420
Qy 421 GGTITGGCAACAGCCAGCGCTAACAGATTAA 456
Db 421 GGTITGGCAACAGCCAGCGCTAACAGATTAA 456

RESULT 9

US-09-543-407-15
Sequence 15, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15

```

?   LENGTH: 456
?   TYPE: DNA
?   ORGANISM: Artificial Sequence
?   FEATURE:
?   OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
?   OTHER INFORMATION: sequence containing the replacement fragment
?   OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
?   US-0543-407-15

```

Query Match	77.2%	Score 352	DB 23	Length 450
Best Local Similarity	85.7%	Pred. No. 1.6e-101		
Matches 391, Conservative	0	Mismatches 65	Indels 0	Gaps 0

QY	ATGAACTTTTAAAGGGGAGCATTCGAGGAAATCGGTTCTGGAGGCTCTGGCT	60
Db	1 ATGAACTTTTAAAGGGGAGCATTCGAGGAAATCGGTTCTGGAGGCTCTGGCT	60
QY	61 GCGCGTGTTCACAATGGGGCGGCGGCTAACTCAATACGGCGGGGAGATGTTCCGGC	120
Db	61 GCGCGTATATGATCAGCTGGTTACCGGTGTTTCCATGAAATGGACATGCAATCCGGC	120
QY	121 CCGGACTCAACGTTGAGCAATTTATCGATACGGTTCGGCTAACGCTGCTCTGCAA	180
Db	121 CCGGACTCAACGTTGAGCAATTTATCGATACGGTTCGGCTAACGCTGCTCTGCAA	180
QY	181 AGCGATGCCCGTAAATCTGAAAAGCAACATTAACCGACGCGTTATGTTAAGCGCGCGAT	240
Db	181 AGCGATGCCCGTAAATCTGAAAAGCAACATTAACCGACGCGTTATGTTAAGCGCGCGAT	240
QY	241 GTAGGCCAGGGTCCGATTAATAGTACTATTGAATGACTCAGAATGGTTTCAGAAATAT	300
Db	241 GTAGGCCAGGGTCCGATTAATAGTACTATTGAATGACTCAGAATGGTTTCAGAAATAT	300
QY	301 GCCACCATCGACAGGTGGAACGCTAAAACTCCGATTTACCTGTCGCGGCAATATGATCAG	360
Db	301 GCCACCATCGACAGGTGGAACGCTAAAACTCCGATTTACCTGTCGCGGCAATGCGGGCT	360
QY	361 CTGGTACCCGTGTGTTAACCATGAAATGGCAATGCAATGCAACGTAATGGTCCGTCAAGTT	420
Db	361 AATTAACGCGCGCTGTGTTATCAGACCGCATCTGATTCACGCTAATATGATCCGTCAAGTT	420
QY	421 GGTGTTGGCAACAAGCCCAACGGCTAACCAAGTATTA	456
Db	421 GGTGTTGGCAACAAGCCCAACGGCTAACCAAGTATTA	456

```

RESULT 10
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

Query Match      76.8%   Score 350.4;   DB 23;   Length 456;
Best Local Similarity 85.5%;   Pred. No. 5,1e-101;

```

Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY	1	ATGAAACTTTTAAAGGTGGCAGCATCTCCACAGCAATCGTAGTTTCTGGACGTGCTGGGCT	60
Db	1	ATGAAACTTTTAAAGTGGCAGCATCTCCACAGCAATCGTAGTTTCTGGCAGTCTCTGGCT	60
QY	61	GCGCTGTTCCCAATGGGGCGCGCGGGTAATCATAAACGCGCGGCAGATAGTCCGGC	120
Db	61	GCGCTGTTCCCAATGGGGCGCGCGGGTAATCATAAACGCGCGGCAGATAGTCCGGAC	120
QY	121	CCGAGCTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACCGCTGCTGTCGCA	180
Db	121	CCGAGCTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACCGCTGCTGCTGCA	180
QY	181	AGCGATGCCGCTAAATCTGAAACGACCATTAACCCAGACGGTTATGTTAACGCGCCGAT	240
Db	181	AGCGATGCCCGTAAATCTGAAACGACCATTAACCCAGACGGTTATGTTAACGCGCCGAT	240
QY	241	GTAAGCCACGGGTGGGGATTAATAGTACTTGAATCTGACATCTCAGATAGTGTTCAGAAATAT	300
Db	241	GTAAGCCACGGGTGGGGATTAATAGTACTTGAATCTGACATCTCAGATAGTGTTCAGAAATAT	300
QY	301	GCCACCATTCGACCAAGTGAGCGTAAACCTCCGATATTACTGTGCGGCAATATGATCAG	360
Db	301	GCAACATGAGACCAAGTGAGCGTAAACCTCCGATATTACTGTGCGGCAATATGCGCGGT	360
QY	361	CTGGTTACCCGTGTTGTTACCATGAAATGSCACATGCAACGGTATAGTGTCAGGTT	420
Db	361	AATAACGCGCGCTGTTAATCAACCGCATCTGATTCACAGCGTATAGTGTCAGGTT	420
QY	421	GGTTTGGCAACAAGCCAGCGGTAAACAGATTAA	456
Db	421	GGTTTGGCAACAAGCCAGCGGTAAACAGATTAA	456

```

RESULT 11
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding P13 from GPe3 of Leishmania major.
US-09-543-407-27

```

	Query Match	76.8%	Score 350.4	DB 23	Length 456
	Best Local Similarity	85.5%	Pred. No. 5.1e-101		
	Matches 390	Conservative 0	Mismatches 66	Indels 0	Gaps 0
QY	1	ATGAACCTTTAAAGTGGCAGCATTCGACATCGTAGTTTCTGGCAGTCTGAGCT	60		
Db	1	ATGAACCTTTAAAGTGGCAGCATTCGACATCGTAGTTTCTGGCAGTCTGAGCT	60		
QY	61	GCGTCGTTCCACAAATGGGCGGCGGGATCATCAAACGGCGGCAATAGTTCGGC	120		
Db	61	GCGTCGTTCCACAAATGGGCGGCGGGATCATCAAACGGCGGCAATAGTTCGGC	120		
QY	121	CGGACCTCAACGTTAGACATTTATCAAGTACGGTTCCGCTAACGCTGGCGTTGCTCTGC	180		


```

Db      121 CCGGACTCAAGCTGAGATTATTCAGTACGGTCCGCTAACGGCCCTGCTCTGCA 180
Qy      181 AGCGATGCCCTTAATCTGAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Db      181 AGCGATGCCCTTAATCTGAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Qy      241 GTAGCCGAGGGTGGCGATTAATGACTATGAACTGAACTGAAATGTTTGAAGAATAT 300
Db      241 TATGATCAGCTGTTTACCCGTTGTTTACCGTAAATGGACATGCAATTCAGAAATAT 300
Qy      301 GCCACATCGACGACGTGAACGCTMAAACTCCGATATTAATGTCGCCCAATATATGAG 360
Db      301 GCCACATCGACGACGTGAACGCTMAAACTCCGATATTAATGTCGCCCAATATGAG 360
Qy      361 CTGGTATCCCGTGTGTTTATCCCATGAATAATGGACATGCAAGCTTAATGTCGAGTT 420
Db      361 AATAACGGCGCGCTGTTTATTCAGACCCGATCTGATTCAGCGTAATGTCGTCAGGTT 420
Qy      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456
Db      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456

```

```

RESULT 12
US-09-543-407-29
; Sequence 29, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PasteSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

```

```

Query Match      76.5%; Score 348.8; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 1.7e-100;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
Qy      1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCCTGAGTCTCTGCT 60
Db      1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCCTGAGTCTCTGCT 60
Qy      61 GCGCTGTTCCAAATGCGGCGCGCGCGGTATCAATACGCGCGCAATGTTCCGCG 120
Db      61 GCGCTGTTCCAAATGCGGCGCGCGCGGTATCAATACGCGCGCAATGTTCCGCG 120
Qy      121 CCGGACTCAAGCTGAGATTATTCAGTACGGTCCGCTAACGCTGCGCTGCTCTGAA 180
Db      121 CCGGACTCAAGCTGAGATTATTCAGTACGGTCCGCTAACGCTGCGCTGCTCTGAA 180
Qy      181 AGCGATGCCCTTAATCTGAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Db      181 AGCGATGCCCTTAATCTGAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Qy      241 GTAGCCGAGGGTGGCGATTAATGACTATGAACTGAACTGAAATGTTTGAAGAATAT 300
Db      241 GTAGCCGAGGGTGGCGATTAATGACTATGAACTGAACTGAAATGTTTGAAGAATAT 300

```

```

Qy      301 GCCACATCGACGACGTGAACGCTMAAACTCCGATATTAATGTCGCCCAATATGAG 360
Db      301 GCCACATCGACGACGTGTTTACCCGTTTACCGTAAATGGACATGCAAGCGCGT 360
Qy      361 CTGGTATCCCGTGTGTTTATCCCATGAATAATGGACATGCAAGCTTAATGTCGAGTT 420
Db      361 AATAACGGCGCGCTGTTTATTCAGACCCGATCTGATTCAGCGTAATGTCGTCAGGTT 420
Qy      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456
Db      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456

```

```

RESULT 13
US-08-233-642A-54
; Sequence 54, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: -
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233.642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; TELEPHONE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEEDANBERRY
; INFORMATION FOR SEQ ID NO: 54:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-642A-54

```

```

Query Match      63.0%; Score 287.2; DB 6; Length 361;
Best Local Similarity 99.0%; Pred. No. 8.7e-81;
Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy      64 GTCTTCACCAATGGGCGCGCGGTATCATACGCGCGCAATAGTTCGGGCGG 123
Db      1 GTCTTACCAACAGTGGGCGCGCGGTATCATACGCGCGCAATAGTTCGGGCGG 60
Qy      124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGGCTTCTCTGCAAGC 183
Db      61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGGCTTCTCTGCAAGC 120

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds

(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456

Sequence: 1 atgaacttttaaaagtgc.....ccacgctaaccgtattaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patents NA New:
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225.6	49.5	675	US-09-741-873C-3	Sequence 3, Appli
2	181.6	39.8	396	US-09-741-873C-1	Sequence 1, Appli
3	33	7.2	3990	US-09-830-230A-629	Sequence 629, App
4	32.4	7.1	150223	US-10-624-149A-1	Sequence 1, Appli
5	31.4	6.9	1588	US-10-767-701-14522	Sequence 14522, A
6	30.6	6.7	7096	US-10-021-698A-892	Sequence 892, App
7	30.6	6.7	196083	US-10-021-698A-892	Sequence 707, App
8	30.2	6.6	2007	US-10-775-972-153	Sequence 153, App
9	30.2	6.6	2148	US-10-775-972-154	Sequence 154, App
10	30	6.6	1355	US-10-045-674A-593	Sequence 593, App
11	29.6	6.5	2159	US-10-767-701-13730	Sequence 13730, A
12	29.6	6.5	24900	US-10-767-701-10736	Sequence 10736, A
13	29.6	6.5	318752	US-10-767-701-10553	Sequence 10553, A
14	29.4	6.4	300	US-10-767-701-28437	Sequence 28437, A
15	29.2	6.4	1790242	US-10-767-701-10805	Sequence 10805, A
16	29	6.4	670	US-10-767-701-15726	Sequence 15726, A
17	29	6.4	861	US-10-767-701-3616	Sequence 3616, App
18	28.8	6.3	1857	PCT-US04-05654-2371	Sequence 2371, App
19	28.6	6.3	1186	PCT-US04-05654-1030	Sequence 1030, App
20	28.6	6.3	4634	US-10-791-799-10	Sequence 10, Appli
21	28.4	6.2	1543	US-10-767-701-1782	Sequence 1782, App
22	28.4	6.2	394468	US-60-548-091-5725	Sequence 5725, App
23	28.2	6.2	596	US-10-767-701-5378	Sequence 5378, App
24	28.2	6.2	610	US-10-767-701-8781	Sequence 8781, App
25	28.2	6.2	798	US-10-100-663-2192	Sequence 2192, App
26	28.2	6.2	1746	PCT-US04-05654-2555	Sequence 2555, App

ALIGNMENTS

27	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
28	28	6.1	714	6	US-10-779-543-23469	Sequence 23469, A
29	27.8	6.1	390	5	US-09-796-592B-3647	Sequence 3647, App
30	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
31	27.6	6.1	201	6	US-10-767-701-28293	Sequence 28293, A
32	27.6	6.1	402	5	US-09-969-034-1464	Sequence 1464, App
33	27.6	6.1	1052	6	US-10-767-701-16629	Sequence 16629, A
34	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
35	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, App
36	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, App
37	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, App
38	27.6	6.1	10251	6	US-10-045-674A-582	Sequence 582, App
39	27.6	6.1	232882	6	US-10-767-701-10679	Sequence 10679, A
40	27.4	6.0	819	6	US-10-417-884A-2244	Sequence 2244, App
41	27.4	6.0	823	6	US-10-767-701-20007	Sequence 20007, A
42	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, App
43	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
44	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, App
45	27.2	6.0	547	6	US-10-767-701-24545	Sequence 24545, A

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 880123-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match
Best Local Similarity 49.5%; Score 225.6; DB 5; Length 675;
Query Similarity 68.4%; Pred. No. 1.4e-66;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGAGCATTTGGACGATCTGTTTGGCAGTCTCTGGCT 60
DB 83 ATGAACCTTTAAAGTGGAGCATTTGGACGATCTGTTTGGCAGTCTCTGGCT 142
QY 61 GGCCTGCTTCAACATGGGGGGGGGGGATATCATTAACGGGGGGGCAATAGTTCGGC 120
DB 143 GGTGTGTCTCTCGTACGGGGGGGGGGGATATCATTAACGGGGGGGCAATAGTTCGGC 202
QY 121 CCGACTCAACGTTAGACATTTTACGATACGTTCCGCTAACGCTGCGTCTCTGCA 180

```
Db 203 CCAAAATTCGAGCTGAACATTTACAGTACGGTGGGCTTACTTGCACCTTGTCTGCAA 262
Qy 181 AGGATGCCCCGTAATCTGAACAGCATTACCCAGAGCGGTTATGCTTAACGGCCGAT 240
Db 263 ACTGATGCCCGGTAACCTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACT 322
Qy 241 GTAGGCGAGGGGTGGGTAATAGTACTGATTTGAATCTGACTGAGATGTTTCAAAATAT 300
Db 323 GTTGTGTCAGGGCTCAAGTACAGCTCAATGATGATGACCAAGTGGCTTCCGTAACAGC 382
Qy 301 GCCACATGCAACGAGTGAACGCTAAATTCGATATGATGCTGCGCCATATGATGAG 360
Db 383 GCTACTCTGATGATGAGTGAACGCAAAATTTCTGAATGACGCTTAAACAGTTCGCTGT 442
Qy 361 CTGGTACCCGCTGTGTCTTACCCATGAAATGACATGCAAGGTAATGCTGCTGAGTT 420
Db 443 GGCACGCTGCTGCACTTACGACGAGCTGATCTAATCTCTCCGTAACGCTGAGT 502
Qy 421 GGTTCGCAACACGCGCAGGCTAACGATTTAA 456
Db 503 GGGTTTGTAAACAGCGGACCGCTCATCTGACTTAA 538
```

RESULT 2

```
US-09-741-873C-1
/ Sequence 1, Application US/09741873C
/ GENERAL INFORMATION:
/ APPLICANT: Olsen, Steffan
/ TITLE OF INVENTION: Fibrinectin Binding Protein As Well As Its Preparation
/ FILE REFERENCE: 012889-084
/ CURRENT APPLICATION NUMBER: US/09/741,873C
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: SE 8801723-1
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: US 08/978,878
/ PRIOR FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: US 07/347,189
/ PRIOR FILING DATE: 1989-05-04
/ PRIOR APPLICATION NUMBER: US 07/789,437
/ PRIOR FILING DATE: 1991-11-06
/ PRIOR APPLICATION NUMBER: US 07/970,846
/ PRIOR FILING DATE: 1992-11-03
/ PRIOR APPLICATION NUMBER: US 08/187,865
/ PRIOR FILING DATE: 1994-01-28
/ PRIOR APPLICATION NUMBER: US 08/318,519
/ PRIOR FILING DATE: 1994-10-05
/ NUMBER OF SEQ ID NOS: 11
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 1
/ LENGTH: 396
/ TYPE: DNA
/ ORGANISM: Escherichia coli
US-09-741-873C-1
```

```
Query Match 39.8%; Score 181.6; DB 5; Length 396;
Best Local Similarity 66.2%; Pred. No. 8.4e-52;
Matches 262; Conservative 0; Mismatches 134; Indels 0; Gaps 0;
```

```
Qy 61 GGCCTGCTCCCAATAGGGCGGGCGGTAATCATTAACGCGCGGCAATGTTCCGCG 120
Db 1 GGTCTTCTCTCTAGTACGGGCGGGGAGTACCAAGTGTGCGGTATATATACCGGC 60
Qy 121 CCGGACTCAACGTTGACATTTATCAATACGTTCCGTTACGCTGCTGCTGCAA 180
Db 61 CCAATTCGAGCTGAACATTTACAGTACGTTGGGGTAACCTTCACCTTCTCTGCAA 120
Qy 181 AGGATGCCCGTAATCTGAACAGCATTACCCAGAGCGGTTATGCTTAACGCGCGAT 240
Db 121 ACTGATGCCCGTAACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACT 180
Qy 241 GTAGGCGAGGGTGGGTAATAGTACTGATTTGAATCTGACTGAGATGTTTCAAAATAT 300
```

```
Db 181 GTTGTGAGGGCTCAGATGACAGCTCAATGATCTGACCCAGGTGCTTGGTAAACAGC 240
Qy 301 GCCACATGCAACGAGTGAACGCTAAATTCGATATGATGCTGCGCCATATGATGAG 360
Db 241 GCTACTCTGATGATGAGTGAACGCAAAATTTCTGAATGACGCTTAAACAGTTCGCTGT 300
Qy 361 CTGGTACCCGCTGTGTCTTACCCATGAAATGACATGCAAGGTAATGCTGCTGAGTT 420
Db 301 GGCACGCTGCTGCACTTACGACGAGCTGATCTAATCTCTCCGTAACGCTGAGT 360
Qy 421 GGTTCGCAACACGCGCAGGCTAACGATTTAA 456
Db 361 GGGTTTGTAAACAGCGGACCGCTCATCTGACTTAA 396
```

RESULT 3

```
US-09-830-230A-629
/ Sequence 629, Application US/09830230A
/ GENERAL INFORMATION:
/ APPLICANT: Human Genome Sciences, Inc.
/ TITLE OF INVENTION: Lyme Disease Vaccines
/ FILE REFERENCE: PB481US
/ CURRENT APPLICATION NUMBER: US/09/830,230A
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: PCT/US98/12718
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/057,483
/ PRIOR FILING DATE: 1997-09-03
/ PRIOR APPLICATION NUMBER: 60/053,344
/ PRIOR FILING DATE: 1997-07-22
/ PRIOR APPLICATION NUMBER: 60/053,377
/ PRIOR FILING DATE: 1997-07-22
/ PRIOR APPLICATION NUMBER: 60/050,359
/ PRIOR FILING DATE: 1997-06-20
/ NUMBER OF SEQ ID NOS: 756
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 629
/ LENGTH: 3990
/ TYPE: DNA
/ ORGANISM: Homo sapiens
```

```
/ FEATURE:
/ LOCATION: (1135)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1143)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1146)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1210)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1247)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1250)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc.feature
/ LOCATION: (1251)
/ OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629
```

Query Match 7.2%; Score 33; DB 5; Length 3990;
Best Local Similarity 47.8%; Pred. No. 0.43;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGGCTGCTCTCCTCAAGCCATGCCGTAATCTGAACGACCATTAACCAG 216
Db 3433 GTTACGGCTGGCGCTGCTGTGAGAGAGATGAGAGAGAGCTGACAGAGCTTCAAAATCCG 3492

Qy 217 AGCGGTTATGCTAACCGCGCCGATGTAGGCCAGGGTGGGATATAGTACTATTGAATG 276
Db 3493 ATTCTGCTGCTATTTGGGAGAGGTAATGAGATGCTGGGATTTTGGTAAGGATGAGATG 3552

Qy 277 ACTGAGATGGTTTCAGAAATATATGACCATGACGACGTAAGCGTAAAGCTCCGAT 336
Db 3553 AAGAGAGATGATCAAGATTGCTGCTGCTATTGAGGGGAGATGAGGATGAGAAAG 3612

Qy 337 ATTACTGTGCGCAATATGAT 357
Db 3613 TTGCTGTGAGAGATATGAT 3613

RESULT 4

US-10-624-149A-1
Sequence 1, Application US/10624149A
GENERAL INFORMATION:
APPLICANT: Neubauer, Antonie
TITLE OF INVENTION: gm-Negative EHV-Mutants without Heterologous Elements
FILE REFERENCE: 1/1372
CURRENT APPLICATION NUMBER: US/10/624,149A
CURRENT FILING DATE: 2003-07-21
PRIOR APPLICATION NUMBER: 60/403,282
PRIOR FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: DE 10233064
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: DE 10317008
PRIOR FILING DATE: 2003-04-11
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Version 3.0
SEQ ID NO 1
LENGTH: 150223
TYPE: DNA
ORGANISM: Equine herpesvirus 1
PUBLICATION INFORMATION:
AUTHORS: Telford, E.A.
AUTHORS: Watson, M.S.
AUTHORS: McBride, K.
AUTHORS: Davison, A.J.
TITLE: The DNA sequence of equine herpesvirus-1
JOURNAL: Virology
VOLUME: 189
ISSUE: 1
PAGES: 304-316
DATE: JUL-1992
DATABASE ACCESSION NUMBER: NC 001491, NCBI
DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 3.6;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 173 CTCTGCAAGGATGCCGTAATCTGAAAGACCATTAACCAAGCGGTTAGTAACG 232
Db 95542 CTGTGTGTAACGCTGGCGCTTGTGATACGGCTCAATTCACAGCATGTTGCTTAACG 95601

Qy 233 GCGCGCATGTAGGCGAGGTCGGATATAGTACTATTGAATGACATGAGATGTTTCA 292
Db 95602 TCAAGCCATGAGGCGAGACCGGATATGTTTCTGTATACCAATCTCTGGGCGCGCTCA 95661

Qy 293 GAATATATGCCCATGACCATGCG 318

Db 95662 GAGAGCTTATTCACAGAGCTGTTG 95687

RESULT 5

US-10-767-701-14522
Sequence 14522, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yinhua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 14522
LENGTH: 1588
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

Query Match 6.9%; Score 31.4; DB 6; Length 1588;
Best Local Similarity 44.5%; Pred. No. 0.99;
Matches 125; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy 40 GTTTCGCAAGTCTCTGCTGCTGCGTGTTCACATGATGGGCGCGCGGTAATCATAC 99
Db 455 GGTGTGGCGGCTTCGGTGGTGGTGTATGAGCGGTGGCGCGGTGCTATGTGTGGC 514

Qy 100 GCGCGCGCAATAGTTCGCGCGCGGACTCAACGTTGACATTTATAGTACGTTCCGCT 159
Db 515 GCGCGTGTATGT 574

Qy 160 AACGCTGCGTGTGCTGCAAGCGATGCCGTAATCTGAAGACCATTAACCAAGC 219
Db 575 GGT 634

Qy 220 GATTATGTAAAGCGCGCGATGTAGCGCAGGTCGCGATATAGTACTATTGAATGACT 279
Db 635 GTAGTAACAGCTTGTGCAAGAGTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 694

Qy 280 CAGATGCTTTCAGAAATATATCCCATGATGACCATGAGAA 320
Db 695 GCTGTAGCTTGTGCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 735

RESULT 6

US-10-021-698A-892
Sequence 892, Application US/10021698A
GENERAL INFORMATION:
APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN BERDRECH, PAUL
APPLICANT: DUPUIS, JOSE
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
APPLICANT: PANDIT, SUNIL
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044051
CURRENT APPLICATION NUMBER: US/10/021,698A
CURRENT FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
PRIOR FILING DATE: 2000-06-14
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: Patentin 2.1
SEQ ID NO 892
LENGTH: 7096
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (7085)..(7087)
OTHER INFORMATION: a, t, c or g
US-10-021-698A-892

Query Match
Best Local Similarity 48.6%; Score 30.6; DB 6; Length 7096;
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 278 CTCGAGATGTTTCAGAAATTAATGCCACATGACGACGAGAAACGCTAAATCTCCGATA 337
DB 2964 CTCCTCAAGGGCCAGAGAAACAATCTTAAGATTAAGTCACTGGCCATTAACAACCTTA 3023
QY 338 TTACTGTGGCCAAATATGATCAGCTGTTACCCGTGTTGTTACCATGAATGCGACATG 397
DB 3024 TTCAGAGAAATTAATCTCTCTGCAAGGCGACCTGTGTGTGAAGCTGAGACCTGTAGT 3083
QY 398 CAAGCTAATGCTGCTGAGTTGTTTGGCAACAGCCGCTAACCG 450
DB 3084 TCAGCAACGAGATCGGCCAAGTGGGGCTGCAGTAACCCCTGGCTTACCG 3136

RESULT 7
US-10-021-698A-707

Sequence 707, Application US/10021698A
GENERAL INFORMATION:
APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN EERDEWEGH, PAUL
APPLICANT: DUPUIS, JOSE
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
APPLICANT: PANDIT, SUNIL
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044US1
CURRENT APPLICATION NUMBER: US/10/021.698A
CURRENT FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
PRIOR FILING DATE: 2000-06-14
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: PatentIn 2.1
SEQ ID NO: 707
LENGTH: 196083
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (256)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (89496)..(8955)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (141214)..(141313)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (182468)..(182567)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (193506)..(193605)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (194799)..(194898)
OTHER INFORMATION: a, t, c or g

FEATURE:
NAME/KEY: modified_base
LOCATION: (195354)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (195460)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (195644)
OTHER INFORMATION: a, t, c or g
US-10-021-698A-707

Query Match
Best Local Similarity 48.6%; Score 30.6; DB 6; Length 196083;
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 278 CTCGAGATGTTTCAGAAATTAATGCCACATGACGACGAGAAACGCTAAATCTCCGATA 337
DB 45914 CTCCTCAAGGGCCAGAGAAACAATCTTAAGATTAAGTCACTGGCCATTAACAACCTTA 45973
QY 338 TTACTGTGGCCAAATATGATCAGCTGTTACCCGTGTTGTTACCATGAATGCGACATG 397
DB 45974 TTCAGAGAAATTAATCTCTCTGCAAGGCGACCTGTGTGTGAAGCTGAGACCTGTAGT 46033
QY 398 CAAGCTAATGCTGCTGAGTTGTTTGGCAACAGCCGCTAACCG 450
DB 46034 TCAGCAACGAGATCGGCCAAGTGGGGCTGCAGTAACCCCTGGCTTACCG 46086

RESULT 8
US-10-775-972-153/c

Sequence 153, Application US/10775972
GENERAL INFORMATION:
APPLICANT: Henderson, Robert A.
APPLICANT: Bangur, Chaitanya S.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.455C21
CURRENT APPLICATION NUMBER: US/10/775.972
CURRENT FILING DATE: 2004-02-10
NUMBER OF SEQ ID NOS: 563
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 153
LENGTH: 2007
TYPE: DNA
ORGANISM: Homo sapiens
US-10-775-972-153

Query Match
Best Local Similarity 45.8%; Score 30.2; DB 6; Length 2007;
Matches 104; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 140 TTATCAGTACGGTTCGCTAAGCGCTGCTGCAACAGATGCCGTAATCTG 199
DB 1416 TTCCCAAGAAATTTCTTAAGAAATGTTATTTTAAAGTAAAGTAAAGT 1357
QY 200 AAACGACATTAACCCAGAGCGTTAAGTAAAGCGCCGAGTAAAGCGAGGCGGATA 259
DB 1356 CACTGGGAATTTTGCAATGAGTTGAGAAACTGAGAGCTGCCCTAGGGAATTTCACTT 1297
QY 260 ATAGTACTATTAAGTCACTCAAGATGTTTCAAGAAATATGCAACCATGACCACTGGA 319
DB 1296 AGGCTTGAATTAATCACTCAAGAGAAATCCAGATTAATCTGCTCCCTCAATAGTGT 1237
QY 320 ACCTAAATCTCCGATATTACTGTGCGCAATATGATAGCTGTT 366
DB 1236 AGGAGAGAGCTCAAGTTCAGAGCTTCTCATTTTACCAACAGTT 1190

RESULT 9

```

US-10-775-972-154/C
: Sequence 154, Application US/10775972
: GENERAL INFORMATION:
: APPLICANT: Henderson, Robert A.
: APPLICANT: Wang, Tongtong
: TITLE OF INVENTION: Chaitanya S.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: FILE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.455C21
: CURRENT APPLICATION NUMBER: US/10/775,972
: NUMBER OF SEQ ID NOS: 563
: SOFTWARE: PasteSeq for Windows Version 4.0
: SEQ ID NO 154
: LENGTH: 2148
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-775-972-154

```

Query Match	6.6%	Score 30.2;	DB 6;	Length 2148;
Best Local Similarity	45.8%	Pred. No. 2.9;		
Matches 104;	Conservative	0;	Mismatches 123;	Indels 0;
			Gaps	0;

QY	140	TTTATCAGTAGGGTTCGCGTAACGTGGCTTGCTGCAAAAGGATCCCGTAAATCTG	199
Db	1549	TTCCCCGAAAATTCTCTCAAAAGATGTTCAATTTTATTAAGTTAAATACTACTAAAGTT	1490
QY	200	AAACGACCATTTACCAGAGCGGTTATGTGAACGGCGCCGATGTAGGCGAGGTCGCGATA	259
Db	1489	CACGTGGGAATTTTGCAATGCAAGTTGAGAAACTGGAGCTGCTACGGGAAATTTCAGCATT	1430
QY	260	ATACTACTATTGAACCTGACTCAGAAATGTTTCAGAAATATGCAACCATGCAACAGTGA	319
Db	1429	AGGCGTTGAATTAATCTACCCAGAGAAATCCCAAGTAAATCTGCGCTCCCATTAATGTGT	1370
QY	320	ACCGTAAAACTCCGATATTACTGTGGCCCAATATGATCACTAGTGT	366
Db	1369	AGGAGAGAGCTCCAAAGTTCGAGAGCTTCTCATTTTACCAAAACAGTT	1323

```

RESULT 10
US-10-045-674A-593
/ Sequence 593, Application US/10045674A
/ GENERAL INFORMATION:
/ APPLICANT: LADNER, ROBERT C.
/ APPLICANT: COHEN, EDWARD H.
/ APPLICANT: NASTRI, HORACIO G.
/ APPLICANT: ROOKEY, KRISTIN L.
/ APPLICANT: HOET, RENE
/ APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
/ TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
/ TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
/ TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
/ FILE REFERENCE: DYA/002 CIP2
/ CURRENT APPLICATION NUMBER: US/10/045,674A
/ CURRENT FILING DATE: 2001-10-25
/ PRIOR APPLICATION NUMBER: 06/198,069
/ PRIOR FILING DATE: 2000-04-17
/ PRIOR APPLICATION NUMBER: 09/837,306
/ PRIOR FILING DATE: 2001-04-17
/ NUMBER OF SEQ ID NOS: 635
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 593
/ LENGTH: 1355
/ TYPE: DNA
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: M13-III
/ OTHER INFORMATION: nucleotide sequence
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(1305)

```

US-10-045-674A-593

Query Match	6.6%;	Score 30;	DB 6;	Length 1355;
Best Local Similarity	25.1%;	Pred. No. 2.7;		
Matches	84;	Conservative	80;	Mismatches 170;
				Indels 0;
				Gaps 0

Qy 57 GCGCGGCGTGTCCAAATGGGGCGGGGGGTAATGAATGAACGGCGGGCAATATGTC 116

Dd 768 KGGTGGTGGTWSYSGYGGGWSYGGTGGTWSYGARGGYGGGWSYGAARGGCG 827

Qy 117 CGGCCCGGACCAACGTTGAGCATTTATCACTACGGTCCCGCTAACGTCGGCTGTCT 176

Dd 828 YGGTWSYSGRSGYGGWSYSGYWSYGGCATTTTGATYAYABABATGGCHAYGCTYA 887

Qy 177 GCAAGCGGATGCCCGTAAATCTGAAGCAACCATTAACCGAGCGGTTATGTAACGGCGC 236

Dd 888 YAAAGSGCATYATGACYABAAVGCYAGABAAYGCRTRCARWSTGAYGCAANAGTYAA 947

Qy 237 CGATGTAGGCCAGGGTGGGATTAATAGTACTATTGAACGTACTAGATGGTTTCAGAAA 296

Dd 948 RYTTGAWSYGTCCGYACMGAYTAAGTGTGCTGATCGAAGYGTTYATVYGGYAGVTWS 1007

Qy 297 TAAATGCCAATCGAACCAAGGAAACGCTAAAACCTCGATTATCTGTGGCCAAATGA 356

Dd 1008 YGGTCTKGCTAAAGYAAVGGMGVACYCGGAGATTTGCGWGYTCKAATTCYCARATGGC 1067

Qy 357 TCAGCTGTTAACCGGTGTTTACCCATGAAATG 390

Dd 1068 YCARGTTCGAGYAGKGAAYAAVWCCCTYATG 1101

RESULT 11
US-10-767-701-13730
; Sequence 13730, Application US/10767701

APPLICANT: Kovalic, David K.
 APPLICANT: Zhou, Yihua
 APPLICANT: Gao, Yongwei
 TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 FILE REFERENCE: 38-21(53535)B
 CURRENT APPLICATION NUMBER: US/10/767,701
 CURRENT FILING DATE: 2004-01-29
 NUMBER OF SEQ ID NOS: 63128
 SEQ ID NO 13730
 LENGTH: 2159
 TYPE: DNA
 ORGANISM: Sorghum bicolor
 FEATURE:
 OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS49140_1
 US-10-767-701-13730

Query Match	6.5%	Score 29.6;	DB 6;	Length 2159;
Best Local Similarity	48.8%;	Pred. No. 4.6;		
Matches 80;	Conservative 0;	Mismatches 84;	Indels 0;	Gaps 0;

Qy	288	TTTCAGAAATAATGCCACCATGACCACTGGAGCTTAAAACTTCGATATTACTGTCCG	347
Db	682	TTACACTATTTCATGACACCGAGGACACGGAAACTGCCAAAAAGATATCTCAAGTTATGA	741
Qy	348	CCAAATGATCACTGTGTTACCCGTTGTTGTTACCATGAAATGGCACATGCAAGCTAAT	407
Db	742	CCAGCAGGGGAGCAAGCCCCCTGTGTGATTTCTCAAAATTTTGAGACCAATGTGTCCGAT	801
Qy	408	GGGCGCTCAGGTGGTTTGGCAACAAGCCACGGCTTAACAGT	451
Db	802	GCGTGGTAGTGTTAAGTTACCCCAAAAGTCAAGAAATCAATGT	845

RESULT 12.
US-10-767-471-10736/C
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:

```
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CL001505
CURRENT APPLICATION NUMBER: US/10/767,471
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10736
LENGTH: 24900
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1) (24900)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736
```

```
Query Match
Best Local Similarity 52.4%; Score 29.6; DB 6; Length 24900;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

```
QY 207 CATTAACCCAGACGGTTATGTAAGCGCGGAGTAGCCAGGCGGTGGAATATAGTAC 266
DB 10048 CAGCATCCTGACTGCTTTAGTGTCTCAAAAGTCTCGGTCAAGTGTGATATAGTGTG 9989
QY 267 TATTGAAGTGAAGTGAAGTGTTCAGAAATATAGCCCATGACCGAGTGAACGCTAA 326
DB 9988 TCCTGAACCCAACTGATGTATGTAATAAATAGTCAATCTGTTGTATATATTTAA 9929
QY 327 AAAC 330
DB 9928 AATC 9925
```

```
RESULT 13
US-10-767-471-10553
; Sequence 10553, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10553
; LENGTH: 318752
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1) (318752)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10553
```

```
Query Match
Best Local Similarity 52.4%; Score 29.6; DB 6; Length 318752;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

```
QY 267 TATTGAAGTGAAGTGAAGTGTTCAGAAATATAGCCCATGACCGAGTGAACGCTAA 326
DB 254839 TGAATACACACCCAGAAACCTTGAGAAATCATGACAGAGATGATATCTCAAA 254898
QY 327 AAACCGATATTAAGTGTGGCAATATGATGAGTGTGTTACCGGTCTTACCATGA 386
DB 254899 ACTTTCGATTTTATGTGGCATATTTGCGATGTGTGAAGGTTTATAGAACTTA 254958
QY 387 AATG 390
DB 254959 AATG 254962
```

```
RESULT 14
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kowalic, David K.
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5353)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 7551688
US-10-767-701-28437
```

```
Query Match
Best Local Similarity 54.1%; Score 29.4; DB 6; Length 300;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
```

```
QY 14 AAGTGCAGCAATTCGAGCAATGTAAGTTCTGGAGTGTCTGCGGTGCTTCAC 73
DB 123 AAGAAAGCCGTTACGAAAGCCGGTGTACGATGCGGCGCGCGGTTACGCGGTCC 182
QY 74 AATGGGCGCGCGGTTATCATTAACGGCGGCGGCAATAGTCCGCCCG 124
DB 183 GTAAAGCGCGCGGTTATGCGGCGCGGCGGCGGTTACGCGGCGCAACCGG 233
```

```
RESULT 15
US-10-767-471-10805/c
; Sequence 10805, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CL001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10805
; LENGTH: 1790242
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1) (1790242)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10805
```

```
Query Match
Best Local Similarity 53.5%; Score 29.2; DB 6; Length 1790242;
Matches 61; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
```

```
QY 266 CTATTGAAGTGAAGTGAAGTGTTCAGAAATATAGCCCATGACCGAGTGAACGCTA 325
DB 369207 CTGTTCTGTATGATGAGATTAATCAAGTAACTGCCCTAAGTCTCTCAAGATCTACT 369148
QY 326 AAACCTCGATATTAAGTGTGGCAATATGATGAGTGTGTTACCGGTGTTA 379
DB 369147 CAAGCTCAGTTATTCGTTAGCTTATATCTTTTCTTACCATTTTAA 369094
```

Search completed: March 17, 2004, 08:25:44
Job time : 73.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 : Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaactcttaaaagtgcg.....ccacgcttaaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_Main:

```
1: /cgn2_6/ptodata/2/pna/US06_PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
```

```
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	23	Sequence 15, Appl
2	403.2	88.4	456	6	Sequence 56, Appl
3	403.2	88.4	456	23	Sequence 19, Appl
4	360	78.9	456	23	Sequence 17, Appl
5	353.6	77.5	456	23	Sequence 11, Appl
6	352	77.2	456	23	Sequence 13, Appl
7	352	77.2	456	23	Sequence 23, Appl
8	352	77.2	456	23	Sequence 25, Appl
9	352	77.2	456	23	Sequence 27, Appl
10	348.8	76.5	456	23	Sequence 29, Appl
11	348.8	76.5	456	23	Sequence 54, Appl
12	345.6	75.8	456	23	Sequence 172, Appl
13	284.2	62.3	361	6	Sequence 172, Appl
14	222.4	48.8	456	23	Sequence 172, Appl
15	222.4	48.8	477	17	Sequence 172, Appl
16	222.4	48.8	477	17	Sequence 172, Appl
17	222.4	48.8	477	17	Sequence 172, Appl
18	222.4	48.8	675	13	Sequence 172, Appl
19	222.4	48.8	675	13	Sequence 172, Appl
20	220	48.2	453	92	Sequence 172, Appl
21	220	48.2	453	92	Sequence 172, Appl
22	180	39.5	396	13	Sequence 172, Appl
23	155.2	34.0	360	31	Sequence 172, Appl
24	64.8	14.2	522	17	Sequence 172, Appl
25	64.8	14.2	522	17	Sequence 172, Appl
26	64.8	14.2	522	17	Sequence 172, Appl
27	51.2	11.2	100	45	Sequence 172, Appl
28	49	10.7	78	23	Sequence 172, Appl
29	48	10.5	48	23	Sequence 172, Appl
30	48	10.5	78	23	Sequence 172, Appl
31	42.4	9.3	100	45	Sequence 172, Appl
32	41.2	9.0	78	23	Sequence 172, Appl
33	40.8	8.9	100	45	Sequence 172, Appl
34	40.6	8.9	78	23	Sequence 172, Appl
35	35.8	7.9	561	17	Sequence 172, Appl
36	35.8	7.9	561	17	Sequence 172, Appl
37	35.8	7.9	561	17	Sequence 172, Appl
38	35.6	7.8	1965	1	Sequence 172, Appl
39	35.6	7.8	1965	1	Sequence 172, Appl
40	35.6	7.8	1965	1	Sequence 172, Appl
41	35.6	7.8	1965	1	Sequence 172, Appl
42	35.4	7.8	3411	1	Sequence 172, Appl
43	35.4	7.8	3411	1	Sequence 172, Appl
44	34.6	7.6	438	31	Sequence 172, Appl
45	34.6	7.6	438	31	Sequence 172, Appl

ALIGNMENTS

RESULT 1
US-09-543-407-15
Sequence 15, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afa
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GPe3 of Leishmania major.
US-09-543-407-15

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.1e-139;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATCGATGTTCTGCGAGTCTGCGCT 60
Db 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATCGATGTTCTGCGAGTCTGCGCT 60
QY 61 GGGCTATATATAGCTGTTTACCGCTGTTTATCCCATGAAATGGACATCATCCGCG 120
Db 61 GGGCTATATATAGCTGTTTACCGCTGTTTATCCCATGAAATGGACATCATCCGCG 120
QY 121 CCGACTCAACGTTGAGCATTTATCAGTACGCTGCGTCAAGCGCTGCTGCGCA 180
Db 121 CCGACTCAACGTTGAGCATTTATCAGTACGCTGCGTCAAGCGCTGCTGCGCA 180
QY 181 AGGATGCCCGTAAATCTGAAACGACATTACCGAGCGTTATGTTAACGCGCGAT 240
Db 181 AGGATGCCCGTAAATCTGAAACGACATTACCGAGCGTTATGTTAACGCGCGAT 240
QY 241 GTAGCCAGGCTGCGATATAGTACTATTTAACTGAACTGCAAGTGTTCAGAAATAT 300
Db 241 GTAGCCAGGCTGCGATATAGTACTATTTAACTGAACTGCAAGTGTTCAGAAATAT 300
QY 301 GCACCATGACCATGAGCAAGCTTAAATCTCCATTTACTGTGCGCAATACGCGCGT 360
Db 301 GCACCATGACCATGAGCAAGCTTAAATCTCCATTTACTGTGCGCAATACGCGCGT 360
QY 361 AATAAGCCGCTGCTTAAATGAGCGCATCTGATTCACGCTTAAATGCTGCGT 420
Db 361 AATAAGCCGCTGCTTAAATGAGCGCATCTGATTCACGCTTAAATGCTGCGT 420
QY 421 GGTTTGGCAACACGCGCATTCACGCTTAAATGCTGCGT 456
Db 421 GGTTTGGCAACACGCGCATTCACGCTTAAATGCTGCGT 456

RESULT 2
US-08-233-642A-56
Sequence 56, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 88.4%; Score 403.2; DB 6; Length 456;
Best Local Similarity 92.8%; Pred. No. 6.1e-122;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
QY 61 GCGCTATGATCAGCTGTTACCCGCTGTTTACCCAGTAATGSCATGATCCGGC 120
DB 61 GCGCTATGATCAGCTGTTACCCGCTGTTTACCCAGTAATGSCATGATCCGGC 120
QY 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
DB 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
QY 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
DB 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCCCAAGGCTGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCCCAAGGCTGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACCGATGGAACGCTAAACTCCGATATTATTCGCGCAATACGGCGCT 360
DB 301 GCCACATCGACCGATGGAACGCTAAACTCCGATATTATTCGCGCAATACGGCGCT 360
QY 361 AATAACGCGCGCTGTTAATCATGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAACGCGCGCTGTTAATCATGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTTTGGCAACAACGCGCAGGCTAACGATATTAA 456
DB 421 GGTTTGGCAACAACGCGCAGGCTAACGATATTAA 456
```

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1

Query Match 88.4%; Score 403.2; DB 23; Length 456;
Best Local Similarity 92.8%; Pred. No. 6.1e-122;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
QY 61 GCGCTATGATCAGCTGTTACCCGCTGTTTACCCAGTAATGSCATGATCCGGC 120
DB 61 GCGCTATGATCAGCTGTTACCCGCTGTTTACCCAGTAATGSCATGATCCGGC 120
QY 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
DB 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
QY 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
DB 121 CCGACTCAACGTTGAGCATTTATCATGATCGGTTCCGTAACGCTGCCCTTCTGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCCCAAGGCTGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCCCAAGGCTGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACCGATGGAACGCTAAACTCCGATATTATTCGCGCAATACGGCGCT 360
DB 301 GCCACATCGACCGATGGAACGCTAAACTCCGATATTATTCGCGCAATACGGCGCT 360
QY 361 AATAACGCGCGCTGTTAATCATGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAACGCGCGCTGTTAATCATGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTTTGGCAACAACGCGCAGGCTAACGATATTAA 456
DB 421 GGTTTGGCAACAACGCGCAGGCTAACGATATTAA 456
```

RESULT 4
US-09-543-407-19
Sequence 19, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding p73 from GP63 of Leishmania major.
US-09-543-407-19

Query Match 78.9%; Score 360; DB 23; Length 456;
Best Local Similarity 86.8%; Pred. No. 1.2e-107;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
QY 61 GCGCTATGATCAGCTGTTACCCGCTGTTTACCCAGTAATGSCATGATCCGGC 120
```

```

Db      61 GGGGTGTTCCAAATGAGGCGCGCGGTATCATTAACGCGCGCATAGTTCCGGC 120
Qy      121 CCGGACTCAAGCTTGAAGCATTTATATCAGTAAGTTCCGCTAACGCTGCTGCTGCA 180
Db      121 CCGGACTCAAGCTTGAAGCATTTATATCAGTAAGTTCCGCTAACGCTGCTGCTGCA 180
Qy      181 AGCGATGCGCGTAAATCTGAACGACATTAACCCAGAGGGTTATGTAACGCGCCGAT 240
Db      181 AGCGATGCGCGTAAATCTGAACGACATTAACCCAGAGGGTTATGTAACGCGCCGAT 240
Qy      241 GTAGGCGGAGGTGCGGATATAGTACTATTGTAAGTACTGACAGATGTTTCAAAATAT 300
Db      241 GTAGGCGGAGGTGCGGATATAGTACTATTGTAAGTACTGACAGATGTTTCAAAATAT 300
Qy      301 GCGACCATGACCAAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Db      301 GCGACCATGACCAAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Qy      361 AATAAGCGCGCGTGGTTATAGACCGCATCTGATTCAGCGTAAATGTTGCTCAGGT 420
Db      361 AATAAGCGCGCGTGGTTATAGACCGCATCTGATTCAGCGTAAATGTTGCTCAGGT 420
Qy      421 GGTITGGCAACACGCGCAAGCTTAAACCTCCGATATTAATGTGCGCCAAATCGCGGT 480
Db      421 GGTITGGCAACACGCGCAAGCTTAAACCTCCGATATTAATGTGCGCCAAATCGCGGT 480

```

RESULT 5

```

US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

```

```

Query Match      77.5%; Score 353.6; DB 23; Length 456;
Best Local Similarity 86.0%; Pred. No. 1.6e-105;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```

```

Qy      1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCTAGTTCTGCGAGTCTGCTGCT 60
Db      1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCTAGTTCTGCGAGTCTGCTGCT 60
Qy      61 GGGGTCTATGATCAGCTGTTTACCGGTGTTTATCCATGAATGGACATGCTCGGC 120
Db      61 GGGGTCTATGATCAGCTGTTTACCGGTGTTTATCCATGAATGGACATGCTCGGC 120
Qy      121 CCGGACTCAAGCTTGAAGCATTTATATCAGTAAGTTCCGCTAACGCTGCTGCTGCA 180
Db      121 CCGGACTCAAGCTTGAAGCATTTATATCAGTAAGTTCCGCTAACGCTGCTGCTGCA 180
Qy      181 AGCGATGCGCGTAAATCTGAACGACATTAACCCAGAGGGTTATGTAACGCGCCGAT 240
Db      181 AGCGATGCGCGTAAATCTGAACGACATTAACCCAGAGGGTTATGTAACGCGCCGAT 240

```

```

Qy      241 GTAGGCGGAGGTGCGGATATAGTACTATTGTAAGTACTGACAGATGTTTCAAAATAT 300
Db      241 GTAGGCGGAGGTGCGGATATAGTACTATTGTAAGTACTGACAGATGTTTCAAAATAT 300
Qy      301 GCGACCATGACCAAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Db      301 GCGACCATGACCAAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Qy      361 AATAAGCGCGCGTGGTTATAGACCGCATCTGATTCAGCGTAAATGTTGCTCAGGT 420
Db      361 AATAAGCGCGCGTGGTTATAGACCGCATCTGATTCAGCGTAAATGTTGCTCAGGT 420
Qy      421 GGTITGGCAACACGCGCAAGCTTAAACCTCCGATATTAATGTGCGCCAAATCGCGGT 480
Db      421 GGTITGGCAACACGCGCAAGCTTAAACCTCCGATATTAATGTGCGCCAAATCGCGGT 480

```

RESULT 6

```

US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11

```

```

Query Match      77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 5.3e-105;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

```

Qy      1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCTAGTTCTGCGAGTCTGCTGCT 60
Db      1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCTAGTTCTGCGAGTCTGCTGCT 60
Qy      61 GGGGTCTATGATCAGCTGTTTACCGGTGTTTATCCATGAATGGACATGCTCGGC 120
Db      61 GGGGTCTATGATCAGCTGTTTACCGGTGTTTATCCATGAATGGACATGCTCGGC 120
Qy      121 CCGGACTCAAGCTTGAAGCATTTATATCAGTAAGTTCCGCTAACGCTGCTGCTGCA 180
Db      121 CCGGACTCAAGCTTGAAGCATTTATATCAGTAAGTTCCGCTAACGCTGCTGCTGCA 180
Qy      181 AGCGATGCGCGTAAATCTGAACGACATTAACCCAGAGGGTTATGTAACGCGCCGAT 240
Db      181 AGCGATGCGCGTAAATCTGAACGACATTAACCCAGAGGGTTATGTAACGCGCCGAT 240
Qy      241 GTAGGCGGAGGTGCGGATATAGTACTATTGTAAGTACTGACAGATGTTTCAAAATAT 300
Db      241 GTAGGCGGAGGTGCGGATATAGTACTATTGTAAGTACTGACAGATGTTTCAAAATAT 300
Qy      301 GCGACCATGACCAAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Db      301 GCGACCATGACCAAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Qy      361 AATAAGCGCGCGTGGTTATAGACCGCATCTGATTCAGCGTAAATGTTGCTCAGGT 420
Db      361 AATAAGCGCGCGTGGTTATAGACCGCATCTGATTCAGCGTAAATGTTGCTCAGGT 420

```

Qy 421 GGTGGGCAACAGCGCGCTTACCAAGATTAA 456
Db 421 GCAATGCAACAGCGCGCTTACCAAGATTAA 456

RESULT 7

US-09-543-407-13
Sequence 13, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-13

Query Match 77.2%; Score 352; DB 23; Length 456;

Best Local Similarity 85.7%; Pred. No. 5.3e-105; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 65;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTCATGATGATGAGCTGTTACCCGTTGTTTACCATGAAATGGCATGATCCGGC 120
Db 61 GGCCTCATGATGATGAGCTGTTACCCGTTGTTTACCATGAAATGGCATGATCCGGC 120
Qy 121 CCGACATCAAGTGGAGATTATGACATGAGTCCGTAACGCTGCGCTGCTCTGCAA 180
Db 121 CCGACATCAAGTGGAGATTATGACATGAGTCCGTAACGCTGCGCTGCTCTGCAA 180
Qy 181 AGCGATGCGCGTAAATCTGAAGACATTAACCGAGAGCGTTATGTAAGCGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAGACATTAACCGAGAGCGTTATGTAAGCGCGCGAT 240
Qy 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTGAGATGTTTCAAGAAATAT 300
Db 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTGAGATGTTTCAAGAAATAT 300
Qy 301 GCCACATCGACGAGTGAAGCGTAAAACTCCGATATTAATGTCGGCCATATAGCGCGT 360
Db 301 GCCACATCGACGAGTGAAGCGTAAAACTCCGATATTAATGTCGGCCATATAGCGCGT 360
Qy 361 AATAAGCGCGCGTGTAAATCAAGACGATCTGATTCAGCGTAATAGTGTGCTGAGTT 420
Db 361 CTGGTACCGCGTGTGTAAATCAAGACGATCTGATTCAGCGTAATAGTGTGCTGAGTT 420
Qy 421 GGTGGGCAACAGCGCGCTTACCAAGATTAA 456
Db 421 GGTGGGCAACAGCGCGCTTACCAAGATTAA 456

RESULT 8

US-09-543-407-23
Sequence 23, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-23

Query Match 77.2%; Score 352; DB 23; Length 456;

Best Local Similarity 85.7%; Pred. No. 5.3e-105; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 65;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTCATGATGATGAGCTGTTACCCGTTGTTTACCATGAAATGGCATGATCCGGC 120
Db 61 GGCCTCATGATGATGAGCTGTTACCCGTTGTTTACCATGAAATGGCATGATCCGGC 120
Qy 121 CCGACATCAAGTGGAGATTATGACATGAGTCCGTAACGCTGCGCTGCTCTGCAA 180
Db 121 CCGACATCAAGTGGAGATTATGACATGAGTCCGTAACGCTGCGCTGCTCTGCAA 180
Qy 181 AGCGATGCGCGTAAATCTGAAGACATTAACCGAGAGCGTTATGTAAGCGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAGACATTAACCGAGAGCGTTATGTAAGCGCGCGAT 240
Qy 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTGAGATGTTTCAAGAAATAT 300
Db 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTGAGATGTTTCAAGAAATAT 300
Qy 301 GCCACATCGACGAGTGAAGCGTAAAACTCCGATATTAATGTCGGCCATATAGCGCGT 360
Db 301 GCCACATCGACGAGTGAAGCGTAAAACTCCGATATTAATGTCGGCCATATAGCGCGT 360
Qy 361 AATAAGCGCGCGTGTAAATCAAGACGATCTGATTCAGCGTAATAGTGTGCTGAGTT 420
Db 361 CATGAATGCGACATGGAATCAAGACGATCTGATTCAGCGTAATAGTGTGCTGAGTT 420
Qy 421 GGTGGGCAACAGCGCGCTTACCAAGATTAA 456
Db 421 GGTGGGCAACAGCGCGCTTACCAAGATTAA 456

RESULT 9

US-09-543-407-25
Sequence 25, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-25

Query Match
Best Local Similarity 77.2%; Score 352; DB 23; Length 456;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTCTGCT 60
QY 61 GGGCTCTATGATCAGCTGTTACCGTGTGTTACCATGAAATGGCAGCATGCTCGGC 120
DB 61 GGGCTCTATGATCAGCTGTTACCGTGTGTTACCATGAAATGGCAGCATGCTCGGC 120
QY 121 CCGGACTCAACGTTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAGCGCGCAT 240
DB 181 CTGGTACCGCTGTTGTTACCATGAAATGGCAGCATGCTGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
DB 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
QY 301 GCCACATGACGACATGGAACGCTTAAACCTCCGATTTACTGTCGCAATAGCGCGT 360
DB 301 GCCACATGACGACATGGAACGCTTAAACCTCCGATTTACTGTCGCAATAGCGCGT 360
QY 361 AATAACGCCGCGCTGTTATGATGACGCGCATCTGATTCGTCGCAATAGCGCGT 420
DB 361 AATAACGCCGCGCTGTTATGATGACGCGCATCTGATTCGTCGCAATAGCGCGT 420
QY 421 GGTTTGGCAACGACGCGCTAACGATTTAA 456
DB 421 GGTTTGGCAACGACGCGCTAACGATTTAA 456
```

RESULT 10
US-09-543-407-21
Sequence 21, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match
Best Local Similarity 76.5%; Score 348.8; DB 23; Length 456;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTCTGCT 60
QY 61 GGGCTCTATGATCAGCTGTTACCGTGTGTTACCATGAAATGGCAGCATGCTCGGC 120
DB 61 GGGCTCTATGATCAGCTGTTACCGTGTGTTACCATGAAATGGCAGCATGCTCGGC 120
QY 121 CCGGACTCAACGTTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
DB 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
QY 301 GCCACATGACGACATGGAACGCTTAAACCTCCGATTTACTGTCGCAATAGCGCGT 360
DB 301 GCCACATGACGACATGGAACGCTTAAACCTCCGATTTACTGTCGCAATAGCGCGT 360
QY 361 AATAACGCCGCGCTGTTATGATGACGCGCATCTGATTCGTCGCAATAGCGCGT 420
DB 361 AATAACGCCGCGCTGTTATGATGACGCGCATCTGATTCGTCGCAATAGCGCGT 420
QY 421 GGTTTGGCAACGACGCGCTAACGATTTAA 456
DB 421 GGTTTGGCAACGACGCGCTAACGATTTAA 456
```

RESULT 11
US-09-543-407-27
Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-27

Query Match
Best Local Similarity 76.5%; Score 348.8; DB 23; Length 456;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTCTGCT 60
QY 61 GGGCTCTATGATCAGCTGTTACCGTGTGTTACCATGAAATGGCAGCATGCTCGGC 120
DB 61 GGGCTCTATGATCAGCTGTTACCGTGTGTTACCATGAAATGGCAGCATGCTCGGC 120
QY 121 CCGGACTCAACGTTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTCTGCA 180
```


QY 230 ACGGCGCGGATGATGAGCGGATGCGGATTAATGACTATTGAACTGACTCAGAAATGTT 289
 DB 167 ACGGCGCGGATGATGAGCGGATGCGGATTAATGACTATTGAACTGACTCAGAAATGTT 226
 QY 290 TCAGAAATTAATGCGGATGAGCGGATGCGGATTAATGACTATTGAACTGACTCAGAAATGTT 349
 DB 227 TCAGAAATTAATGCGGATGAGCGGATGCGGATTAATGACTATTGAACTGACTCAGAAATGTT 286
 QY 350 AATACGGCGGATTAATGAGCGGATGCGGATTAATGACTATTGAACTGACTCAGAAATGTT 398
 DB 287 AATACGGCGGATTAATGAGCGGATGCGGATTAATGACTATTGAACTGACTCAGAAATGTT 335

RESULT 14 US-09-543-407-3

Sequence 3, Application US/09543407
 GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 APPLICANT: Kay, William W.
 TITLE OF INVENTION: BACTERIAL PIMERIAL SYSTEM FOR
 TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
 FILE REFERENCE: 920043,406
 CURRENT APPLICATION NUMBER: US/09/543,407
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 3
 LENGTH: 456
 TYPE: DNA
 ORGANISM: E. coli
 US-09-543-407-3

Query Match 48.8%; Score 222.4; DB 23; Length 456;
 Best Local Similarity 68.0%; Pred. No. 3,9e-62;
 Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTTCGACCAATCGTAGTTTGGCAGTGGCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGCAGCATTTCGACCAATCGTAGTTTGGCAGTGGCTGGCT 60
 QY 61 GCGCTTATGATCAGCTGCTTACCGCTTGTTCACCAATGAGCAATGATCGGC 120
 DB 61 GCGCTTATGATCAGCTGCTTACCGCTTGTTCACCAATGAGCAATGATCGGC 120
 QY 121 CCGAATCTGAGTGAACATTTACAGTACCGTGGGTAACCTGCTGCTGCTGCA 180
 DB 121 CCGAATCTGAGTGAACATTTACAGTACCGTGGGTAACCTGCTGCTGCTGCA 180
 QY 181 AGCGATCGGCTTAATCTGAAACGACATTCACCAAGCGGTTATGAGTAAAGCGCGCAT 240
 DB 181 AGCGATCGGCTTAATCTGAAACGACATTCACCAAGCGGTTATGAGTAAAGCGCGCAT 240
 QY 241 GTAGCGCAGGCTGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 241 GTAGCGCAGGCTGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 QY 301 GCGACATCGACGATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 360
 DB 301 GCGACATCGACGATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 360
 QY 361 AATACCGCGGCTGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 420
 DB 361 AATACCGCGGCTGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 420
 QY 421 GGTGTTGCAACAGCGCAGCGCTTAACAGTAAATGAA 456
 DB 421 GGTGTTGCAACAGCGCAGCGCTTAACAGTAAATGAA 456

RESULT 15
 US-09-252-691-172

Sequence 172, Application US/09252691B
 GENERAL INFORMATION:
 APPLICANT: Keith G. Weinstock et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
 TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196,135
 CURRENT APPLICATION NUMBER: US/09/252,691B
 NUMBER OF SEQ ID NOS: 11324
 SEQ ID NO 172
 LENGTH: 477
 TYPE: DNA
 ORGANISM: Enterobacter cloacae
 US-09-252-691-172

Query Match 48.8%; Score 222.4; DB 17; Length 477;
 Best Local Similarity 68.0%; Pred. No. 4e-62;
 Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTTCGACCAATCGTAGTTTGGCAGTGGCTGGCT 60
 DB 22 ATGAACTTTTAAAGTGGCAGCATTTCGACCAATCGTAGTTTGGCAGTGGCTGGCT 81
 QY 61 GCGCTTATGATCAGCTGCTTACCGCTTGTTCACCAATGAGCAATGATCGGC 120
 DB 82 GCGCTTATGATCAGCTGCTTACCGCTTGTTCACCAATGAGCAATGATCGGC 141
 QY 121 CCGAATCTGAGTGAACATTTACAGTACCGTGGGTAACCTGCTGCTGCTGCA 180
 DB 142 CCGAATCTGAGTGAACATTTACAGTACCGTGGGTAACCTGCTGCTGCTGCA 201
 QY 181 AGCGATCGGCTTAATCTGAAACGACATTCACCAAGCGGTTATGAGTAAAGCGCGCAT 240
 DB 202 AGCGATCGGCTTAATCTGAAACGACATTCACCAAGCGGTTATGAGTAAAGCGCGCAT 261
 QY 241 GTAGCGCAGGCTGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 300
 DB 262 GTAGCGCAGGCTGAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 321
 QY 301 GCGACATCGACGATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 360
 DB 322 GCGACATCGACGATGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 381
 QY 361 AATACCGCGGCTGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 420
 DB 382 AATACCGCGGCTGAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 441
 QY 421 GGTGTTGCAACAGCGCAGCGCTTAACAGTAAATGAA 456
 DB 442 GGTGTTGCAACAGCGCAGCGCTTAACAGTAAATGAA 477

Search completed: March 18, 2004, 02:47:01
 Job time : 2924.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds

(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456
Sequence: 1 atgaacttttaaaagtgc.....ccacgctaccagctatcaa 456

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending Patents NA New: *
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq: *
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq: *
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq: *
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq: *
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq: *
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq: *
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	222.4	48.8	675	US-09-741-873C-3	Sequence 3, Appli
2	180	39.5	5	US-09-741-873C-1	Sequence 1, Appli
3	33.4	7.3	3990	US-09-830-230A-629	Sequence 629, App
4	32.4	7.1	150223	US-10-624-149A-1	Sequence 1, Appli
5	30.8	6.8	1575	US-10-788-792-27	Sequence 27, Appli
6	29.6	6.5	24900	US-10-767-471-10736	Sequence 10736, A
7	29.4	6.4	1400	US-60-545-213-2135	Sequence 2135, Ap
8	29.4	6.4	1400	US-60-545-213-2135	Sequence 2135, Ap
9	29.4	6.4	1400	US-60-545-213-6406	Sequence 6406, Ap
10	29.4	6.4	1400	US-60-545-213-6407	Sequence 6407, Ap
11	29.2	6.4	249	US-10-771-241-117	Sequence 117, App
12	29.2	6.4	757	US-10-771-241-29	Sequence 29, Appli
13	28.8	6.3	83009	US-10-417-375A-143	Sequence 143, Appli
14	28.4	6.2	2511	PCT-US04-05654-2123	Sequence 123, Ap
15	28.4	6.2	20043	US-10-021-698A-869	Sequence 869, App
16	28.4	6.2	394468	US-60-548-091-5725	Sequence 5725, A
17	28.2	6.2	201	US-60-548-091-22688	Sequence 22688, A
18	28.2	6.2	1295	US-10-100-683-5480	Sequence 5480, Ap
19	28.2	6.2	3416	US-10-100-683-3397	Sequence 3397, Ap
20	28.2	6.2	8206	US-10-021-698A-3529	Sequence 3529, Ap
21	28	6.1	2007	US-10-775-972-153	Sequence 153, App
22	28	6.1	2148	US-10-775-972-154	Sequence 154, App
23	27.8	6.1	639	US-10-767-701-7904	Sequence 7904, Ap
24	27.8	6.1	1422	US-10-767-795-7135	Sequence 7135, Ap
25	27.8	6.1	2342	US-10-451-467A-277	Sequence 277, App
26	27.8	6.1	2895	US-10-767-701-14458	Sequence 14458, A

27	27.8	6.1	18802	US-10-767-471-10574	Sequence 10574, A
28	27.8	6.1	32372	US-10-767-471-10611	Sequence 10611, A
29	27.8	6.1	112618	US-10-767-471-10615	Sequence 10615, A
30	27.6	6.1	201	US-10-767-471-28293	Sequence 28293, A
31	27.6	6.1	600	US-60-545-213-2020	Sequence 2020, Ap
32	27.6	6.1	600	US-60-545-213-6292	Sequence 6292, Ap
33	27.6	6.1	1498	US-10-100-683-5108	Sequence 5108, Ap
34	27.6	6.1	1523	US-10-100-683-750	Sequence 750, App
35	27.6	6.1	1662	US-10-100-683-5106	Sequence 5106, App
36	27.6	6.1	2571	US-09-999-183A-4	Sequence 4, Appli
37	27.6	6.1	3035	US-10-786-892-279	Sequence 279, App
38	27.6	6.1	232882	US-10-767-471-10679	Sequence 10679, A
39	27.4	6.0	201	US-10-767-471-11899	Sequence 11899, A
40	27.4	6.0	868	US-10-767-701-167	Sequence 167, App
41	27.4	6.0	318752	US-10-767-471-10553	Sequence 10553, A
42	27.2	6.0	959	US-10-767-701-1446	Sequence 1446, Ap
43	27.2	6.0	998	US-09-461-537A-3	Sequence 3, Appli
44	27	5.9	511	US-10-767-701-18763	Sequence 18763, A
45	27	5.9	987	US-10-100-683-3090	Sequence 3090, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Oleen, Arne
TITLE OF INVENTION: Ribonectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978, 878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347, 189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789, 437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970, 846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187, 865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318, 519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 48.8%; Score 222.4; DB 5; Length 675;
Best Local Similarity 68.0%; Pred. No. 8.7e-67;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGCGAGCTCTGGCT 60
DB 83 ATGAACCTTTAAAGTGGCAGCATTCGACGATTCGAGCTCTGCGAGCTCTGGCA 142
QY 61 GCGCTCATGATGCGCTGTTACCCGCTGTTTACCATGAAATGGCAGCATCGCGCC 120
DB 143 GGTGTTCTTCCTCGTAGCGGGGGGGGGGTACCAAGCGTGGCGGTAAATAGCGGC 202
QY 121 CCGAGCTCAACGTTGAGCATTTTACGATGAGCTTCCGCTAACGCTGCGCTGCTGCA 180

```

Db      203 CCAATTCGAGCTGAACATTACCACTAGCGGTGGCGGTAACCTGCACTTGCTCGCA 262
Qy      181 AGCGATGCCCGTAATCTGAACACGACATTACCGAGCGGTAATGTAACGGCCGAT 240
Db      263 ACTATGCCCCCTAATCTGACTTACTATTAACCAAGATGGCGCGGTAATGTTGCAAT 322
Qy      241 GTAGCGCCAGGTTGGGATTAATAGTACTATTAAGTAACTACTCAGAAATGTTTGAATAT 300
Db      323 GTTGTCAGGCTGAGTACAGTACAGTCAATGATCTTACCAAGTGGCTTGATTAACGC 382
Qy      301 GCCACCATGCAACGATGGAACGCTAAACCTCGATATATTCTGTGGCAATAGCGCGCT 360
Db      383 GCTACTCTTATCATAGTGAACCGCAAAATTCGAATGACGTTAAACAGTTCGGTGT 442
Qy      361 AATAAGCCCGCGCTGTTATATAGACCGCATCTGATTCCAGGTAATGTTGCTCAGATT 420
Db      443 GGCACGCGTCTGCACTTACCAAGTGCATCTAATCTCCGTCACAGTACGATGAT 502
Qy      421 GGTTCGGACACACCGCCAGGCTTAACAGTATTAA 456
Db      503 GGCTTTGTGAACACGCGACCGCTCATGACTACTAA 538

```

RESULT 2

```

US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

```

Query Match 39.5%; Score 180; DB 5; Length 396;
 Best Local Similarity 70.6%; Pred. No. 2.7e-52;
 Matches 240; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

Qy      117 CGGCCGAGCACTGTTAGCATTTATGATCGGTTCCGTAACGCTGGCTGCT 176
Db      57 CGGCCCAATTCGATGAACTTTACAGTACGCTGCGGTAATCTGCACTTGCTCT 116
Qy      177 GCAAAAGGATGCCGTAATCTGAAGAGCACTTACCCAGAGCGGTTATGTTAAACGGCGC 236
Db      117 GCAAAAGGATGCCGTAATCTGAACTTATCCAGAGATGGCGCGGTAATGAGTC 176
Qy      237 CGATGAGCGCAGGTTGCGGATTAATGATCTATTGAAGTCACTGAAATGGTTTCAAAA 236
Db      177 AGATGTTGTCAGGCTGATGATGACAGCTCATGATCGATCCAGCTGCTTCGTTAA 236
Qy      297 TAATGCACCATGACAGTGAACGCTAAATCTCCGATTTTACTGTGGCCCAATACGG 356

```

```

Db      237 CAGCGTACTCTTGTATCATGTGAACGGCAAAATTTGAATGACGTTAAACAGTTCCG 296
Qy      357 CGGTAAATAACGCGCGCTGTGTTAATCAGACCGCATCTGATTCAGGTTAATGTTGCTCA 416
Db      297 TGTGTGCAACGCTGCTGCAAGTTGACAGATCTGATCTCTCCGTCAACGTTACTCA 356
Qy      417 GGTGTTTGTGGCAACACGCGCAAGGCTTAACAGTATTAA 456
Db      357 GGTTCGTTGTGTAACACGAGACCGCTCATGACTACTAA 396

```

RESULT 3

```

US-09-830-230A-629
; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 629
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1210)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1244)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1247)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1250)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1251)
; OTHER INFORMATION: n equals a,t,g, or c
; OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

```

Query Match

Best Local Similarity 7.3%; Score 33.4; DB 5; Length 3990;
 Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: Patent version 3.2
; SEQ ID NO 2134
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2134

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCCCGCATGTAGCCGAGGTGCGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATGAGCTTTCAGTACCGGTGACCGAGTGGGGTGATCGGTTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATATGCAACATGACCGAGTGAAGCGTAAAACT 331
DB 377 TTGGGGTTCCTCCATCCACTGCTTGAAGAGCCAGAGACCATCTCTCCACACTCTGGAAT 318
QY 332 CCGATATTACTGTGCGGCAATACGCGCGTATATACCCCGCGTGTATATAGACCGCAT 391
DB 317 CCGTTGAGAGAGCGGAGTGTGATGAGCAATTAACCTCGTGTACTGTCATCAGGTTTCT 258
QY 392 CTGATTCACGCGTATGTCGTCAGATGGTTGTTTGGCAACAACCGCAAGGTTAACCG 450
DB 257 GGGATTCCTCCAGTGTGTGATTAATTTCTGAAGTGAATGAAGAGGGGTGACTCG 199

RESULT 8
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: Patent version 3.2
; SEQ ID NO 2135
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2135

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCCCGCATGTAGCCGAGGTGCGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATGAGCTTTCAGTACCGGTGACCGAGTGGGGTGATCGGTTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATATGCAACATGACCGAGTGAAGCGTAAAACT 331
DB 377 TTGGGGTTCCTCCATCCACTGCTTGAAGAGCCAGAGACCATCTCTCCACACTCTGGAAT 318
QY 332 CCGATATTACTGTGCGGCAATACGCGCGTATATACCCCGCGTGTATATAGACCGCAT 391
DB 317 CCGTTGAGAGAGCGGAGTGTGATGAGCAATTAACCTCGTGTACTGTCATCAGGTTTCT 258

QY 392 CTGATTCACGCGTATGTCGTCAGATGGTTTGGCAACAACCGCAAGGTTAACCG 450
DB 257 GGGATTCCTCCAGTGTGTGATTAATTTCTGAAGTGAATGAAGAGGGGTGACTCG 199

RESULT 9
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: Patent version 3.2
; SEQ ID NO 6406
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6406

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCCCGCATGTAGCCGAGGTGCGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATGAGCTTTCAGTACCGGTGACCGAGTGGGGTGATCGGTTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATATGCAACATGACCGAGTGAAGCGTAAAACT 331
DB 377 TTGGGGTTCCTCCATCCACTGCTTGAAGAGCCAGAGACCATCTCTCCACACTCTGGAAT 318
QY 332 CCGATATTACTGTGCGGCAATACGCGCGTATATACCCCGCGTGTATATAGACCGCAT 391
DB 317 CCGTTGAGAGAGCGGAGTGTGATGAGCAATTAACCTCGTGTACTGTCATCAGGTTTCT 258
QY 392 CTGATTCACGCGTATGTCGTCAGATGGTTTGGCAACAACCGCAAGGTTAACCG 450
DB 257 GGGATTCCTCCAGTGTGTGATTAATTTCTGAAGTGAATGAAGAGGGGTGACTCG 199

RESULT 10
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: Patent version 3.2
; SEQ ID NO 6407
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6407

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCCCGCATGTAGCCGAGGTGCGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATGAGCTTTCAGTACCGGTGACCGAGTGGGGTGATCGGTTATTA 378

Page 5

APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc J
APPLICANT: Reuber, T. Lynne
APPLICANT: Keddie, James
APPLICANT: Dubell III, Arnold N
APPLICANT: Pineda, Omaria
APPLICANT: Repetti, Peter
APPLICANT: Century, Karen
APPLICANT: Guttersen, Neal
APPLICANT: Yu, Guo-Liang
APPLICANT: Broun, Pierre E
APPLICANT: Kumamoto, Roderick W
APPLICANT: Pilgrim, Marsha L
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
FILE REFERENCE: MBI-0047 PCT
CURRENT APPLICATION NUMBER: PCT/US04/05654
PRIOR FILING DATE: 2004-03-03
PRIOR APPLICATION NUMBER: 10/374,780
PRIOR FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: 10/374,780
PRIOR FILING DATE: 2003-09-30
NUMBER OF SEQ ID NOS: 2950
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2123
LENGTH: 2511
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G1548 Predicted polypeptide sequence is paralogous to G390, G391
PCT-US04-05654-2123

Query Match

Best Local Similarity 56.4%; Score 28.4; DB 1; Length 2511;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 21 AGCATTGCGAGCATGTAAGTTCTGCGAGTCTGCGCTGCGCTGTAATGATCAGCTGT 80
DB 1470 AGCAGCAGCAGTAAGTAGAGGCTTGTAGTCCCGAGTGGAGATTGGAGGGCAGGT 1529
QY 81 TACCCGTGTGTTTACCATGAAATGCGACATGCA 114
DB 1530 TAACTTCACCTGCTGCTACATGATGAGCATGAA 1563

RESULT 15

US-10-021-698A-869/C
Sequence 869, Application US/10021698A
GENERAL INFORMATION:
APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN EERDEWEGH, PAUL
APPLICANT: DUPUIS, JOSEE
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044US1
CURRENT APPLICATION NUMBER: US/10/021,698A
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: PatentIn 2.1
SEQ ID NO 869
LENGTH: 20043
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (6103)
OTHER INFORMATION: a, t, c or g

FEATURE:
NAME/KEY: modified_base
LOCATION: (20008)
OTHER INFORMATION: a, t, c or g
US-10-021-698A-869

Query Match

Best Local Similarity 56.4%; Score 28.4; DB 6; Length 20043;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 9 TTAAAGTGGCAGCATTTGCGAATCGTAAGTTCTGCGAGTCTGCGCTGCTCTTA 68
DB 10361 TCTAATCTGTAATCTGCGCTGCTATGTTGGCTCGGCTGTCCACAGCTGGGGCTG 10302
QY 69 TGATCAGCTGGTTACCCGTTGTTTACCCATGAA 102
DB 10301 GGAAGAGAGTGTGCTGCTGTGGAATGCTCGGGA 10268

Search completed: March 17, 2004, 08:25:46
Job time : 71.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaactttaaagtcg.....ccacgcctacacagratraa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Pending Patents NA Main:

1: /cgn2_6/ptodata/2/pna/US0955A_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US0956A_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US0957A_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US0958A_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US0959A_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US0960A_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US0961A_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US0962A_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US0963A_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US0964A_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US0965A_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US0966A_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US0967A_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US0968A_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US0969A_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US0970A_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US0971A_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US0972A_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US0973A_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US0974A_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US0975A_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US0976A_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US0977A_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US0978A_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US0979A_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US0980A_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US0981A_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US0982A_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US0983A_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US0984A_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US0985A_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US0986A_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US0987A_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US0988A_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US0989A_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US0990A_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US0991A_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US0992A_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US0993A_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US0994A_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US0995A_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US0996A_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US0997A_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US1008A_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US1009A_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US1010A_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US1011A_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US1012A_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US1013A_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US1014A_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US1015A_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US1016A_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US1017A_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US1018A_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US1019A_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US1020A_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US1021A_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US1022A_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US1023A_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US1024A_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US1025A_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US1026A_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US1027A_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US1028A_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US1029A_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US1030A_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US1031A_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US1032A_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US1033A_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US1034A_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US1035A_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US1036A_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US1037A_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US1038A_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US1039A_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US1040A_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US1041A_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US1042A_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US1043A_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US1044A_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US1045A_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US1046A_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US1047A_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US1048A_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US1049A_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US1050A_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US1051A_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US1052A_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US1053A_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US1054A_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US1055A_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US1056A_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US1057A_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US1058A_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US1059A_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US1060A_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US1061A_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US1062A_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US1063A_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US1064A_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US1065A_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US1066A_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US1067A_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US1068A_COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US1069A_COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US1070A_COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US1071A_COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US1072A_COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US1073A_COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US1074A_COMB.seq.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	23 US-09-543-407-17	Sequence 17, Appl
2	406.4	89.1	456	6 US-08-233-642A-56	Sequence 56, Appl
3	406.4	89.1	456	23 US-09-543-407-1	Sequence 1, Appl
4	363.2	79.6	456	23 US-09-543-407-19	Sequence 19, Appl
5	355.2	77.9	456	23 US-09-543-407-11	Sequence 11, Appl
6	355.2	77.9	456	23 US-09-543-407-13	Sequence 13, Appl
7	355.2	77.9	456	23 US-09-543-407-23	Sequence 23, Appl
8	353.6	77.5	456	23 US-09-543-407-25	Sequence 25, Appl
9	353.6	77.5	456	23 US-09-543-407-15	Sequence 15, Appl
10	352	77.2	456	23 US-09-543-407-21	Sequence 21, Appl
11	348.8	76.5	456	23 US-09-543-407-27	Sequence 27, Appl
12	282.2	61.9	456	23 US-08-233-642A-54	Sequence 54, Appl
13	230.4	50.5	456	23 US-08-978-878-3	Sequence 3, Appl
14	230.4	50.5	456	23 US-08-978-878-3	Sequence 3, Appl
15	230.4	50.5	456	23 US-09-741-873B-1	Sequence 1, Appl
16	230.4	50.5	456	23 US-09-741-873B-1	Sequence 1, Appl
17	228	50.0	453	92 US-60-444-371-1	Sequence 1, Appl
18	228	50.0	453	92 US-60-444-371-1	Sequence 1, Appl
19	219.2	48.1	477	17 US-09-252-691C-172	Sequence 172, App
20	219.2	48.1	477	17 US-09-252-691C-172	Sequence 172, App
21	219.2	48.1	477	17 US-09-252-691C-172	Sequence 172, App
22	186.4	40.9	396	13 US-08-978-878-1	Sequence 1, Appl
23	161.6	35.4	360	31 US-09-741-873B-1	Sequence 1, Appl
24	64.8	14.2	522	17 US-09-252-691C-170	Sequence 170, App
25	64.8	14.2	522	17 US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	17 US-09-252-691C-170	Sequence 170, App
27	51.2	11.2	100	45 US-10-417-886-170	Sequence 170, App
28	50	11.0	78	23 US-09-543-407-52	Sequence 52, Appl
29	50	11.0	78	23 US-09-543-407-52	Sequence 52, Appl
30	48	10.5	48	23 US-09-543-407-53	Sequence 53, Appl
31	42.4	9.3	100	45 US-10-417-886-170	Sequence 170, App
32	42.4	9.3	100	45 US-10-417-886-170	Sequence 170, App
33	40.8	8.9	100	45 US-10-417-886-170	Sequence 170, App
34	40.6	8.9	100	45 US-10-417-886-170	Sequence 170, App
35	39.4	8.6	4762	62 US-09-543-407-48	Sequence 48, Appl
36	39.4	8.6	4762	62 US-09-543-407-48	Sequence 48, Appl
37	39.4	8.6	4762	62 US-09-543-407-48	Sequence 48, Appl
38	39.4	8.6	2242716	33 US-09-806-866A-13	Sequence 13, Appl
39	39.2	8.6	2242716	33 US-09-806-866A-13	Sequence 13, Appl
40	37	8.1	14652	33 US-10-018-470A-1	Sequence 1, Appl
41	37	8.1	14652	33 US-10-018-470A-1	Sequence 1, Appl
42	37	8.1	2242716	33 US-09-806-866A-30	Sequence 30, Appl
43	36.4	8.0	561	17 US-09-252-691C-171	Sequence 171, App
44	36.4	8.0	561	17 US-09-252-691C-171	Sequence 171, App
45	36.4	8.0	561	17 US-09-252-691C-171	Sequence 171, App

ALIGNMENTS

RESULT 1
US-09-543-407-17
Sequence 17, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 17
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afa
OTHER INFORMATION: Sequence containing the replacement fragment
US-09-543-407-17
encoding PT3 from Gp63 of Leishmania major.

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.2e-133;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAACTTTTAAAGTGGAGCATTCGAGCATGTGTTCTGCGAGTCTGCT	60
DB	1	ATGAACTTTTAAAGTGGAGCATTCGAGCATGTGTTCTGCGAGTCTGCT	60
QY	61	GAGCTGCTTCAATGAGGCGGCGGCTTAATCATACGCGGCAATGATGCGCG	120
DB	61	GAGCTGCTTCAATGAGGCGGCGGCTTAATCATACGCGGCAATGATGCGCG	120
QY	121	CCGACTATATAGCTGTGTTACCCGTTGTTACCAATGAAATGACATGCTGCA	180
DB	121	CCGACTATATAGCTGTGTTACCCGTTGTTACCAATGAAATGACATGCTGCA	180
QY	181	AGCGATGCGGTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAGCGCG	240
DB	181	AGCGATGCGGTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAGCGCG	240
QY	241	GTTAGGCGAGGTCGATTAATGTTAATGAACTGATGATGATGTTTCAAAATTA	300
DB	241	GTTAGGCGAGGTCGATTAATGTTAATGAACTGATGATGATGTTTCAAAATTA	300
QY	301	GCCACCATGACACAGTGAAGCTTAAATCTGTCGCAATGACGCGCT	360
DB	301	GCCACCATGACACAGTGAAGCTTAAATCTGTCGCAATGACGCGCT	360
QY	361	ATAACGCGGCTGTTAATGACAGCGCATGATTCAGAGGTAATGAGTGTCAAGTT	420
DB	361	ATAACGCGGCTGTTAATGACAGCGCATGATTCAGAGGTAATGAGTGTCAAGTT	420
QY	421	GGTTTGGCAACAACGCGGCTAATACGATTTA	456
DB	421	GGTTTGGCAACAACGCGGCTAATACGATTTA	456

RESULT 2
US-08-233-642A-56
Sequence 56, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
BASED VACCINES
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua


```

1  REGISTRATION NUMBER: 35,570
2  REFERENCE/DOCKET NUMBER: 920043.403C3
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: (206) 622-4900
5  TELEFAX: (206) 682-6031
6  TELEEX: 3723816 SPEEDANBERRY
7  INFORMATION FOR SEQ ID NO: 56:
8  SEQUENCE CHARACTERISTICS:
9  LENGTH: 456 base pairs
10 TYPE: nucleic acid
11 STRANDEDNESS: single
12 TOPOLOGY: linear
13 FEATURE:
14 NAME/KEY: CDS
15 LOCATION: 1..456
16 US-08-233-642A-56

```

Query Match	89.1%	Score 406.4	DB 6	Length 456
Best Local Similarity	93.2%	Pred. No. 1.2e-117		
Matches 425; Conservative	0	Mismatches 31	Indels 0	Gaps 0

Qy	ATGAAACTTTTAAAAATGGCAGCATTCCGACAAATCGTAGTTCTGGCAGTGTCTGACT	60
Db	1 ATGAAACTTTTAAAAATGGCAGCATTCCGACAAATCGTAGTTCTGGCAGTGTCTGACT	60
Qy	GGGTGCTTCCAAATGGGGCGGCGGGTATCTAAACGGCGGCGCAATAGTTCGGCG	120
Db	61 GGGTGTGTTCCAAATGGGGCGGCGGGTATCTAAACGGCGGCGCAATAGTTCGGCG	120
Qy	CCGACTATGATCAGCTGGTTACCGGTGTGTTAACCGAATAGGACATGACTCA	180
Db	121 CCGACTATCAGCTGGTTACCGGTGTGTTAACCGAATAGGACATGACTCA	180
Qy	AGCGATGCCGTAAATCTGAAAAGACATTAACCGACGGGTTATGGTAAAGGCGCGAT	240
Db	181 AGCGATGCCGTAAATCTGAAAAGACATTAACCGACGGGTTATGGTAAAGGCGCGAT	240
Qy	GTAGGCGCAGGGTCCGATATAGTACTATGAACTGACTCAGAATGGTTTCAGAAATAT	300
Db	241 GTAGGCGCAGGGTCCGATATAGTACTATGAACTGACTCAGAATGGTTTCAGAAATAT	300
Qy	GCCACCATGACACGAGGAAACGCTAAAACTCCGATATTACTGTCCGCCAATACGGCGT	360
Db	301 GCCACCATGACACGAGGAAACGCTAAAACTCCGATATTACTGTCCGCCAATACGGCGT	360
Qy	AATTAACGCGCGCTGGTTATCAGACCGCATCTGATTCAGACGTAATGGTCGTCAGTT	420
Db	361 AATTAACGCGCGCTGGTTATCAGACCGCATCTGATTCAGACGTAATGGTCGTCAGTT	420
Qy	GGTTTGGCAACAAGCCACGGCTAAACCAAGTTAA	456
Db	421 GGTTTGGCAACAAGCCACGGCTAAACCAAGTTAA	456

```

RESULT 3
US-09-543-407-1
; Sequence 1 Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FINGERAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO. 1
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Salmonella enteritidis

```

US-09-543-407-1

Query Match	89.1%;	Score 406.4;	DB 23;	Length 456;
Best Local Similarity	-93.2%;	Pred. NO. 1.2e-117;		
Matches 425;	Conservative 0;	Mismatches 31;	Indels 0;	Gaps 0;

Qy	1	ATGAACCTTTAAAGTGGCAGCATTTGCAGCAATCGTAAGTTCTGGCAGAGCTCTGGCT	60
Db	1	ATGAACCTTTTAAAGTGGCAGCATTTGCAGCAATCGTAAGTTCTGGCAGAGCTCTGGCT	60
Qy	61	GGCGTGTTCACAAATGGGGCGGCGGCGGTATATCAACGGCGCGGCAATAGTTCCGGC	120
Db	61	GGCGTGTTCACCAATGGGGCGGCGGCGGTATATCAACGGCGCGGCAATAGTTCCGGC	120
Qy	121	CCGGACTATGATCAGCTGGTATCCCGTGTGTGTAACCATGAAATGGCAATGCACTGCA	180
Db	121	CCGGACTCAACGTTATGACATTTATCAGTACGGTTCGGTAAAGCTGGCTGCTGCA	180
Qy	181	AGCGATGCCCTTAATCTGAAACGACACTTACCGACAGCGGTTATGGTAAACGGCGCGAT	240
Db	181	AGCGATGCCCTTAATCTGAAACGACACTTACCGACAGCGGTTATGGTAAACGGCGCGAT	240
Qy	241	GTATGGCCAGGGTGGCGGATATATGTATCTATTTGAACCTGACTCAAAATGGTTTCAGAAATAT	300
Db	241	GTATGGCCAGGGTGGCGGATATATGTATCTATTTGAACCTGACTCAAAATGGTTTCAGAAATAT	300
Qy	301	GGCACCATCGACAGTAGGAAACGCTTAATAATCTCCGATTAATCTGGCGCAATACGGCGGT	360
Db	301	GGCACCATCGACAGTAGGAAACGCTTAATAATCTCCGATTAATCTGGCGCAATACGGCGGT	360
Qy	361	AATTAACGCGCGCTGGTTAATCAAGCCGATCTGATTTCAAGCGTAAATGGTGGCTCAGGTT	420
Db	361	AATTAACGCGCGCTGGTTAATCAAGCCGATCTGATTTCAAGCGTAAATGGTGGCTCAGGTT	420
Qy	421	GGTTTGGCAACAAAGCAAGCGGTAAACAGATATTA	456
Db	421	GGTTTGGCAACAAAGCAAGCGGTAAACAGATATTA	456

```

RESULT 4
US-09-543-407-19
? Sequence 19, Application US/09543407
? GENERAL INFORMATION:
? APPLICANT: White, Aaron P.
? APPLICANT: Doran, James L.
? APPLICANT: Collinson, S. Karen
? APPLICANT: Kay, William W.
? TITLE OF INVENTION: BACTERIAL IMMUNAL SYSTEM FOR
? TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
? FILE REFERENCE: 920043.406
? CURRENT APPLICATION NUMBER: US/09/543,407
? CURRENT FILING DATE: 2000-04-05
? NUMBER OF SEQ ID NOS: 59
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 19
? LENGTH: 456
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
? OTHER INFORMATION: sequence containing the replacement fragment
? US-09-543-407-19
?

```

Query Match	79.6%	Score 363.2	DB 23	Length 456
Best Local Similarity	87.3%	Pred. No. 66-104		
Matches 398	Conservative 0	Mismatches 58	Indels 0	Gaps 0
QY	1	ATGAAACCTTTAAAGGCGACGACATTCGCGACGATCGTAGTTCTTGGCGAGCGCTCTGGCT	60	
	1	ATGAAACCTTTAAAGGCGACGACATTCGCGACGACATGTAAGTTCTTGGCGAGCTCTGGCT	60	
DB				
QY	61	GAGCGTCGTTCCACAAATGAGGCGCGCGCGGTAAATCATTAACGGCGCGGACATAGTTCCGAGC	120	

US-09-543-407-11
RESULT 5
US-09-543-407-11

ID NO 11
 LENGTH: 456
 TYPE: DNA
 for windows Version 4.0

Query Match	Best Local Similarity	Matches	Score	Replacement fragment
77.9%	80	392	100	From GP63 of Leishmania major.



sequence 13, Application
GENERAL INFORMATION: US/09543407
APPLICANT: white
APPLICANT: white

Query Match	Score
Best Local Similarity	77.9%
Matches 393	Score 355

Qy 421 GATTGGCAACACCGCGCTAACGATATTA 456
Db 421 GATTGGCAACACCGCGCTAACGATATTA 456

RESULT 7

US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-23

Query Match 77.9%; Score 355.2; DB 23; Length 456;

Best Local Similarity 86.2%; Pred. No. 2.1e-101;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCTGAGTCTGAGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCTGAGTCTGAGCT 60
Qy 61 GCGCTGTTCCACATGGGGGCGGCGGTATCATTAACGGCGGCAATGTTCCGGC 120
Db 61 GCGCTGTTCCACATGGGGGCGGCGGTATCATTAACGGCGGCAATGTTCCGGC 120
Qy 121 CCGGACTATGATCACTGCTTACCGGTGTGTAACCCATGAATGGCATGCACTGAA 180
Db 121 CCGGACTATGATGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Qy 181 AGCGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTATGTAACGGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTATGTAACGGCGCGAT 240
Qy 241 GTAGCGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGCGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Qy 301 GCCACCATCGACGAGTGAAGCTAAATCTCCATATTACTGTCGGCGCATAGCGCGT 360
Db 301 GCCACCATCGACGAGTGAAGCTAAATCTCCATATTACTGTCGGCGCATAGCGCGT 360
Qy 361 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATTCACAGTAATGGTCCGTCAGGT 420
Db 361 CATGAATGGCAATGCAATCAAGCCGATCTGATTCACAGTAATGGTCCGTCAGGT 420
Qy 421 GATTGGCAACACCGCGCTAACGATATTA 456
Db 421 GATTGGCAACACCGCGCTAACGATATTA 456

RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-25

Query Match 77.9%; Score 355.2; DB 23; Length 456;

Best Local Similarity 86.2%; Pred. No. 2.1e-101;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCTGAGTCTGAGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCTGAGTCTGAGCT 60
Qy 61 GCGCTGTTCCACATGGGGGCGGCGGTATCATTAACGGCGGCAATGTTCCGGC 120
Db 61 GCGCTGTTCCACATGGGGGCGGCGGTATCATTAACGGCGGCAATGTTCCGGC 120
Qy 121 CCGGACTATGATCACTGCTTACCGGTGTGTAACCCATGAATGGCATGCACTGAA 180
Db 121 CCGGACTATGATGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Qy 181 AGCGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTATGTAACGGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTATGTAACGGCGCGAT 240
Qy 241 GTAGCGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGCGCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Qy 301 GCCACCATCGACGAGTGAAGCTAAATCTCCATATTACTGTCGGCGCATAGCGCGT 360
Db 301 GCCACCATCGACGAGTGAAGCTAAATCTCCATATTACTGTCGGCGCATAGCGCGT 360
Qy 361 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATTCACAGTAATGGTCCGTCAGGT 420
Db 361 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATTCACAGTAATGGTCCGTCAGGT 420
Qy 421 GATTGGCAACACCGCGCTAACGATATTA 456
Db 421 GATTGGCAACACCGCGCTAACGATATTA 456

RESULT 9

US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15

LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 US-09-543-407-15

Query Match
 Best Local Similarity 77.5%; Score 353.6; DB 23; Length 456;
 Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGTAATTTCTGGAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGTAATTTCTGGAGTGTCTGGCT 60
QY 61 GGGGTGCTTCCAAATGGGGCGGGCGGTATCATACGGGGGCAATAGTTCGGC 120
DB 61 GGGGTGCTTCCAAATGGGGCGGGCGGTATCATACGGGGGCAATAGTTCGGC 120
QY 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTATCCCAATGCAATGCAATGCAAT 180
DB 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTATCCCAATGCAATGCAATGCAAT 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACGAGCGGTTATGGTAAACGGCGCAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACGAGCGGTTATGGTAAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATATGATGATGAACTGACATGCAATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATATGATGATGAACTGACATGCAATGTTTCAAAATAT 300
QY 301 GCCACATCGACACGCTGAAACGCTAAATCTCCGATATTAGTGGGCAATACGGCGCT 360
DB 301 GCCACATCGACACGCTGAAACGCTAAATCTCCGATATTAGTGGGCAATACGGCGCT 360
QY 361 AATAACCGCGCGCTGTTATATCAGACCGCATCTGATTCAGGTTATGTCGTCAGATT 420
DB 361 AATAACCGCGCGCTGTTATATCAGACCGCATCTGATTCAGGTTATGTCGTCAGATT 420
QY 421 GGTGTTGGCAACAAACCGCGCTAACCGATATTA 456
DB 421 GGTGTTGGCAACAAACCGCGCTAACCGATATTA 456

```

RESULT 10
 US-09-543-407-21
 Sequence 21, Application US/09543407
 GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
 FILE REFERENCE: 920043.406
 CURRENT APPLICATION NUMBER: US/09/543.407
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 21
 LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 US-09-543-407-21

Query Match
 Best Local Similarity 77.2%; Score 352; DB 23; Length 456;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGTAATTTCTGGAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGTAATTTCTGGAGTGTCTGGCT 60
QY 61 GGGGTGCTTCCAAATGGGGCGGGCGGTATCATACGGGGGCAATAGTTCGGC 120
DB 61 GGGGTGCTTCCAAATGGGGCGGGCGGTATCATACGGGGGCAATAGTTCGGC 120
QY 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTATCCCAATGCAATGCAATGCAAT 180
DB 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTATCCCAATGCAATGCAATGCAAT 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACGAGCGGTTATGGTAAACGGCGCAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACGAGCGGTTATGGTAAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATATGATGATGAACTGACATGCAATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATATGATGATGAACTGACATGCAATGTTTCAAAATAT 300
QY 301 GCCACATCGACACGCTGAAACGCTAAATCTCCGATATTAGTGGGCAATACGGCGCT 360
DB 301 GCCACATCGACACGCTGAAACGCTAAATCTCCGATATTAGTGGGCAATACGGCGCT 360
QY 361 AATAACCGCGCGCTGTTATATCAGACCGCATCTGATTCAGGTTATGTCGTCAGATT 420
DB 361 AATAACCGCGCGCTGTTATATCAGACCGCATCTGATTCAGGTTATGTCGTCAGATT 420
QY 421 GGTGTTGGCAACAAACCGCGCTAACCGATATTA 456
DB 421 GGTGTTGGCAACAAACCGCGCTAACCGATATTA 456

```

RESULT 11
 US-09-543-407-27
 Sequence 27, Application US/09543407
 GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
 FILE REFERENCE: 920043.406
 CURRENT APPLICATION NUMBER: US/09/543.407
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 27
 LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 US-09-543-407-27

Query Match
 Best Local Similarity 77.2%; Score 352; DB 23; Length 456;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGTAATTTCTGGAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGTAATTTCTGGAGTGTCTGGCT 60
QY 61 GGGGTGCTTCCAAATGGGGCGGGCGGTATCATACGGGGGCAATAGTTCGGC 120
DB 61 GGGGTGCTTCCAAATGGGGCGGGCGGTATCATACGGGGGCAATAGTTCGGC 120
QY 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTATCCCAATGCAATGCAATGCAAT 180

```

DB 121 CCGGACTGAGGTTAGCATTTATAGTACGGTCCGCTAACGGCTGCTCTGCA 180
QY 181 AGCGATGCCGCTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTAAACGGCCGAT 240
DB 181 AGCGATGCCGCTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTAAACGGCCGAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATAGTACTATGACGACTCAGAAAGTTTCAGAAATAT 300
DB 241 TATGATGACGTGGTTACCCGCTGTGTATCCCATGAAAGGACATGCAATTCAGAAATAT 300
QY 301 GCCACCATGACAGTGGAAACGCTTAAAACTCCGATATTATCTGCGCCAAATACGGCCGT 360
DB 301 GCCACCATGACAGTGGAAACGCTTAAAACTCCGATATTATCTGCGCCAAATACGGCCGT 360
QY 361 AATAACGCCGCGCTGTTAATACAGCCGCACTGATTCACAGGTAATGTTGCTCAGGTT 420
DB 361 AATAACGCCGCGCTGTTAATACAGCCGCACTGATTCACAGGTAATGTTGCTCAGGTT 420
QY 421 GGTTCGCAACAACGCCGCTAACGATTTAA 456
DB 421 GGTTCGCAACAACGCCGCTAACGATTTAA 456

RESULT 12

US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-29

Query Match 76.5%; Score 348.8; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 2.2e-99;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGACCAATCGAGTTCTGCGAGTCTGAGCT 60
DB 1 ATGAAACTTTTAAAGTGGCAGCATTCGACCAATCGAGTTCTGCGAGTCTGAGCT 60
QY 61 GCGCTGTTCCAAATGCGGCGCGCGCGGCTAATCAACGCGCGCGCAATAGTTCCGGC 120
DB 61 GCGCTGTTCCAAATGCGGCGCGCGCGGCTAATCAACGCGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTATGATCAAGCTGTTACCGGTTGTTACCCATGAAATGSCATGACCTGCA 180
DB 121 CCGGACTATGATCAAGCTGTTACCGGTTGTTACCCATGAAATGSCATGACCTGCA 180
QY 181 AGCGATGCCGCTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTAAACGGCCGAT 240
DB 181 AGCGATGCCGCTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTAAACGGCCGAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATAGTACTATGACGACTCAGAAAGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATAGTACTATGACGACTCAGAAAGTTTCAGAAATAT 300

QY 301 GCCACCATGACAGTGGAAACGCTTAAAACTCCGATATTATCTGCGCCAAATACGGCCGT 360
DB 301 GCCACCATGACAGTGGAAACGCTTAAAACTCCGATATTATCTGCGCCAAATACGGCCGT 360
QY 361 AATAACGCCGCGCTGTTAATACAGCCGCACTGATTCACAGGTAATGTTGCTCAGGTT 420
DB 361 AATAACGCCGCGCTGTTAATACAGCCGCACTGATTCACAGGTAATGTTGCTCAGGTT 420
QY 421 GGTTCGCAACAACGCCGCTAACGATTTAA 456
DB 421 GGTTCGCAACAACGCCGCTAACGATTTAA 456

RESULT 13

US-08-233-642A-54
Sequence 54, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-642A-54

Query Match 61.9%; Score 282.2; DB 6; Length 361;
Best Local Similarity 90.1%; Pred. No. 2.8e-78;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 64 GTCGTTCCCAATAGGGCGGCGGCGGCTAATCATTAACGGCGGCAATAGTTCGGCCG 123
DB 1 CTCCTACCAAGTGGGCGGCGGCGGCTAATCATTAACGGCGGCAATAGTTCGGCCG 60
QY 124 GACTATGATCAGTGTGTTACCGGTGTTACCCATGAAATGSCATGACCTGCAAGC 183
DB 61 GACTCAACGTTGACATTTATCACTAGGTTCCGCTAACGCTGCTGCTCAAGC 120

QY 184 GATGCCGTAATCTGAAACGACATTCACAGAGCGGTATGTGTAACGGCGCCGATGTA 243
DB 121 GATGCCGTAATCTGAAACGACATTCACAGAGCGGTATGTGTAACGGCGCCGATGTA 180
QY 244 GGCCGAGGTGGGATATAGTACTGTAAGTGAAGTCAAGTAAAGTTTCAAGAAATATGCC 303
DB 181 GGCCGAGGTGGGATATAGTACTGTAAGTGAAGTCAAGTAAAGTTTCAAGAAATATGCC 240
QY 304 ACCATCCGACGATGGAACGCTTAAATCCGATTAATCTGTCGCCCAATAGCGCGGTAT 363
DB 241 ACCATCCGACGATGGAACGCTTAAATCCGATTAATCTGTCGCCCAATAGCGCGGTAT 300
QY 364 AACCCCGCGTGTATATCAGACCGCATCTGATTC 398
DB 301 AACCCCGCGTGTATATCAGACCGCATCTGATTC 335

RESULT 14

US-09-543-407-3
Sequence 3, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Dorian, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 456
TYPE: DNA
ORGANISM: E. coli
US-09-543-407-3

Query Match 50.5%; Score 230.4; DB 23; Length 456;
Best Local Similarity 69.1%; Pred. No. 8.2e-62;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGGAGCATTCGACGATTCGATGTTCTGGCAGTCTTGCT 60
DB 1 ATGAACCTTTTAAAGTGGGAGCATTCGACGATTCGATGTTCTGGCAGTCTTGCT 60
QY 61 GGCGTCTTCAACATGGGCGCGCGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 120
DB 61 GGCGTCTTCAACATGGGCGCGCGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTATGATCGAGTGTACCCGATGTTTACCCATGAAATGGACATGACAGCA 180
DB 121 CCGGACTATGATCGAGTGTACCCGATGTTTACCCATGAAATGGACATGACAGCA 180
QY 181 AGCGATGCCGCTAAATCTGAACGACATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
DB 181 AGCGATGCCGCTAAATCTGAACGACATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
QY 241 GTAGCGAGGTGGGATATAGTACTGTAAGTGAAGTCAAGTAAAGTTTCAAGAAATAT 300
DB 241 GTAGCGAGGTGGGATATAGTACTGTAAGTGAAGTCAAGTAAAGTTTCAAGAAATAT 300
QY 301 GCCACATCGACGATGGAACGCTTAAATCCGATTAATCTGTCGCCCAATAGCGCGGT 360
DB 301 GCCACATCGACGATGGAACGCTTAAATCCGATTAATCTGTCGCCCAATAGCGCGGT 360
QY 361 AATTAAGCCGCGGTGTATATCAGACCGCATCTGATTCAGCGTAAATGTTGCGTCA 420
DB 361 AATTAAGCCGCGGTGTATATCAGACCGCATCTGATTCAGCGTAAATGTTGCGTCA 420
QY 421 GGTGTTGGCAACAGCGCAAGCGTAAACGATATTA 456
DB 421 GGTGTTGGCAACAGCGCAAGCGTAAACGATATTA 456

RESULT 15

US-08-978-878-3
Sequence 3, Application US/08978878
GENERAL INFORMATION:
APPLICANT: NORMARK, Staffan
APPLICANT: OLSEN, Arne
TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
FILE REFERENCE: 012889-081
CURRENT APPLICATION NUMBER: US/08/978,878
CURRENT FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: SE 8801723-1
EARLIER FILING DATE: 1988-05-06
EARLIER APPLICATION NUMBER: US 07/347,189
EARLIER FILING DATE: 1989-05-04
EARLIER APPLICATION NUMBER: US 07/789,437
EARLIER FILING DATE: 1991-11-06
EARLIER APPLICATION NUMBER: US 07/970,846
EARLIER FILING DATE: 1992-11-03
EARLIER APPLICATION NUMBER: US 08/187,865
EARLIER FILING DATE: 1994-01-28
EARLIER APPLICATION NUMBER: US 08/318,519
EARLIER FILING DATE: 1994-10-05
EARLIER APPLICATION NUMBER: US 08/495,959
EARLIER FILING DATE: 1995-06-28
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-08-978-878-3

Query Match 50.5%; Score 230.4; DB 13; Length 675;
Best Local Similarity 69.1%; Pred. No. 9.4e-62;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGGAGCATTCGACGATTCGATGTTCTGGCAGTCTTGCT 60
DB 83 ATGAACCTTTTAAAGTGGGAGCATTCGACGATTCGATGTTCTGGCAGTCTTGCT 142
QY 61 GGCGTCTTCAACATGGGCGCGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 120
DB 143 GGCGTCTTCAACATGGGCGCGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 202
QY 121 CCGGACTATGATCGAGTGTACCCGATGTTTACCCATGAAATGGACATGACAGCA 180
DB 203 CCGGACTATGATCGAGTGTACCCGATGTTTACCCATGAAATGGACATGACAGCA 262
QY 181 AGCGATGCCGCTAAATCTGAACGACATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
DB 263 AGCGATGCCGCTAAATCTGAACGACATTAACGAGCGGTTATGTTAAGCGCGCGAT 322
QY 241 GTAGCGAGGTGGGATATAGTACTGTAAGTGAAGTCAAGTAAAGTTTCAAGAAATAT 300
DB 323 GTAGCGAGGTGGGATATAGTACTGTAAGTGAAGTCAAGTAAAGTTTCAAGAAATAT 382
QY 301 GCCACATCGACGATGGAACGCTTAAATCCGATTAATCTGTCGCCCAATAGCGCGGT 360
DB 383 GCCACATCGACGATGGAACGCTTAAATCCGATTAATCTGTCGCCCAATAGCGCGGT 442
QY 361 AATTAAGCCGCGGTGTATATCAGACCGCATCTGATTCAGCGTAAATGTTGCGTCA 420
DB 443 AATTAAGCCGCGGTGTATATCAGACCGCATCTGATTCAGCGTAAATGTTGCGTCA 502
QY 421 GGTGTTGGCAACAGCGCAAGCGTAAACGATATTA 456
DB 503 GGTGTTGGCAACAGCGCAAGCGTAAACGATATTA 456

Thu Mar 18 12:28:01 2004

us-09-543-407-17.rnp

Page 9

Search completed: March 18, 2004, 02:47:06
Job time : 2927.24 secs

THIS PAGE IS BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaacttttaaaatgagc.....ccacgcttaaccagctatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Pending Patents NA New*

1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230.4	50.5	675	5	US-09-741-873C-3
2	186.4	40.9	396	5	US-09-741-873C-1
3	34.2	7.5	1355	6	US-10-045-674A-593
4	31.4	6.9	435	6	US-10-021-658A-1216
5	31.4	6.9	2007	6	US-10-775-972-153
6	31.4	6.9	2148	6	US-10-775-972-154
7	31.2	6.8	9278	6	US-10-021-658A-796
8	31	6.8	1099	6	US-10-767-701-8827
9	31	6.8	150223	6	US-10-624-149A-1
10	30.8	6.8	1575	6	US-10-788-782-21
11	30.4	6.7	983	6	US-10-767-701-1287
12	30.2	6.6	889	6	US-10-767-795-6329
13	30	6.6	1203	6	US-10-767-701-12486
14	29.8	6.5	1228	6	US-10-767-701-6233
15	29.6	6.5	656	6	US-10-767-701-20611
16	29.6	6.5	24900	6	US-10-767-701-10736
17	29.6	6.5	186449	6	US-10-021-658A-715
18	29.4	6.4	300	6	US-10-767-701-28437
19	29.4	6.4	1400	7	US-10-767-701-17781
20	29.4	6.4	1400	7	US-60-545-213-2135
21	29.4	6.4	1400	7	US-60-545-213-2135
22	29.4	6.4	1400	7	US-60-545-213-6406
23	29.4	6.4	1400	7	US-60-545-213-6407
24	29.4	6.4	145597	6	US-10-624-149A-2
25	29.4	6.4	145597	6	US-10-624-149A-2
26	29.2	6.4	249	6	US-10-771-241-117

27	29.2	6.4	757	6	US-10-771-241-29	Sequence 29, Appl
28	29	6.4	594	6	US-10-767-701-15041	Sequence 15041, A
29	29	6.4	1002	1	PCT-US04-05854-859	Sequence 859, App
30	29	6.4	1052	6	US-10-767-701-12607	Sequence 12607, A
31	29	6.4	2600	6	US-10-451-467A-39	Sequence 39, Appl
32	28.8	6.3	959	6	US-10-767-701-1446	Sequence 1446, Ap
33	28.8	6.3	83009	6	US-10-417-375A-143	Sequence 143, App
34	28.6	6.3	592	6	US-10-767-701-32979	Sequence 22979, A
35	28.6	6.3	749	6	US-10-767-701-12516	Sequence 12516, A
36	28.6	6.3	1186	1	PCT-US04-05654-1030	Sequence 1030, Ap
37	28.4	6.2	378	6	US-10-417-884A-1059	Sequence 1059, Ap
38	28.4	6.2	394468	7	US-60-548-091-5725	Sequence 5725, Ap
39	28.2	6.2	201	7	US-60-548-091-32688	Sequence 22688, A
40	28.2	6.2	610	6	US-10-767-701-8781	Sequence 8781, Ap
41	28.2	6.2	1295	6	US-10-100-683-5480	Sequence 5480, Ap
42	28.2	6.2	3416	6	US-10-100-683-3397	Sequence 3397, Ap
43	28.2	6.2	3776	7	US-60-548-091-34	Sequence 34, Appl
44	28.2	6.2	3828	7	US-60-548-091-38	Sequence 38, Appl
45	28.2	6.2	8206	6	US-10-021-658A-3529	Sequence 3529, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Oleen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US/09741,873C
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 50.5%; Score 230.4; DB 5; Length 675;
Best Local Similarity 69.1%; Pred. No. 1.1e-68;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGAGCAATTCGAGTTCGAGTCTTGCT 60
DB 83 ATGAACCTTTAAAGTGGAGCAATTCGAGTTCGAGTCTTGCT 142
QY 61 GGCCTGCTTCAATGAGGCGGCGGCTTAATTAACGCGGCGCAATGCTCCG 120
DB 143 GGTGTGTCTTCTGTCGCGGCGGCGGCTTAATTAACGCGGCGCAATG 202
QY 121 CCGACTATGATGACGTGTTACCCGCTGTTGTTACCATGAAATGACATGCACTGCA 180

Db	203	CCAAATTCGAGCTGAACATTTCACGAGTACGGGCGGGTAACTCGACCTTGCTCTGCA	252
OY	181	AGCGATGCCGCTAAATCTGAAAGCAACCATTCACCGAGGGTTATGTAACGGCGCCGAT	240
Db	263	ACTATATCCCGTAACTCTGACTTGATATTAACCAAGCATGGGCGGGTAATGGTGCAGAT	322
OY	241	GTAGCGCAGGTCGCGATTAATGTAATCTATTGAACTGACTCAAAATGTTTCAGAAATAAT	300
Db	323	GTGGTGAGGGCTCAATGACAGCTCAATCGATTCGACCAACGTCGCTTCGGTAACAGC	382
OY	301	GCCACCATCGACCAGTGGAACGCTAAAACTCCGATATTACTGTGGCCATATACGGCGG	360
Db	383	GCTACTCTTGATATAGTGAACGGCAAAATTCGAAATGAAGTTAAACAGTTCCGGTGT	442
OY	361	AATTAACGCGCGCTGTTATATCAGACCGCATCTGATTCACGCGTAATAGTGTGCTCAGTT	420
Db	443	GGCAACGGTGCTGACAGTTGACCAAGCTCACTTAATCTTCCTCGTCAACGTGACTCAGTT	502
OY	421	GGTTTGGGAACAAGCCCAACGGCTAACCAAGATTTAA	456
Db	503	GGCTTTGTTAACAACGACCGGCTCATATGAGTACTTAA	538

```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741, 873C
CURRENT FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: SE 8801723-1
PRIORITY FILING DATE: 1998-05-06
PRIORITY APPLICATION NUMBER: US 08/978, 878
PRIORITY FILING DATE: 1997-11-26
PRIORITY APPLICATION NUMBER: US 07/347,189
PRIORITY FILING DATE: 1989-05-04
PRIORITY APPLICATION NUMBER: US 07/789,437
PRIORITY FILING DATE: 1991-11-06
PRIORITY APPLICATION NUMBER: US 07/970,846
PRIORITY FILING DATE: 1992-11-03
PRIORITY APPLICATION NUMBER: US 08/197,865
PRIORITY FILING DATE: 1994-01-28
PRIORITY APPLICATION NUMBER: US 08/318,519
PRIORITY FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
; US-09-741-873C-1

```

Query Match	40.9%	Score 186.4	DB 5	Length 396
Best Local Similarity	66.9%	Pred. No. 8.2e-54		
Matches	265	Conservative 0	Mismatches 131	Indels 0
			Gaps 0	
QY	61	GGCGTGTTCACAAATGGGCGGCGCGCGTATCATAAAGCGCGCGCAATAGTTCCGGC	120	
Db	1	GGGTGTGTTCTCCACATACCGCGCGCGCGTACACACGTTGTGGCGGTAAATAGCGGC	60	
QY	121	CGGACATATGATCAGCTGTTATCCCGTGTGTATACCAATGGCAATGCAATGCATCTGCA	180	
Db	61	CCAAATTCTGAGCTGAACATTTCACAGTACGGTGGCGGTACTCTGCACTTCTGTGCA	120	
QY	181	AGCGATGCCCGGTAAATCTGAAACGACACATTAACCAAGAGCGGTTATGGTAAAGCGCGCGAT	240	
Db	121	ACTGATGCCCGGTAACTCTGACTTACCTATTAACCAACATAGGCGCGGTAAATGGTGCAGAT	180	
QY	241	GTAGGCCAGGTCGCGGATTAATAGTACTATTTGAACCTGACTCAGAAATGTTTCAGAAATAT	300	

[illegible]

```

RESULT 3
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENÉ
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DYAK/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; PRIORITY FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; IS-10-045-674A-593

```

Query Match	7.5%	Score 34.2	DB 6	Length 1355
Best Local Similarity	26.9%	Fred. No. 0.094		
Matches	90	Conservative	76	Mismatches 169
				Indels 0
				Gaps 0
QY	20	CAGCATTGCGACCAATCGTAGTTTCTG6CAGTGTCTGCTGCGCGTGTCCACAAATGG	79	
Db	755	CMCCCKETAAAYCAKCGYGGYGGYGGYGGMGYSYSGYGGYGYMYSGARRGYCGWG	814	
QY	80	GCGCGCGCGGTATATATAAGCGCGCGCGCAATAGTTCCGCCCGGACCTATGATAGCTGG	139	
Db	815	GWMSYGARBMGMYGGYMSYSGRGGYGGYMSYGGYMSYGGYGAATYYGATYATVGAHARA	874	
QY	140	TTAACCCGTTGTTTACCATGAATAATGGCACATGCACTGTGCAAGCGATCCCGTAAATCTG	199	
Db	875	TGGCAAAAGCAAAAYARSGCYATGACYARAAAYCAYGARAAAYGRCITRCARMSWG	934	
QY	200	AAACGACCATTAACCCAGAGCGGTTATAGTADAGCGCGCATTTAGCCAGGTCGGGATA	259	
Db	935	AVGCTAANAGTAATYMTAGWATSYGTCGCTACMCATYATAGTGTCGATVCAAYGAYTTVA	994	
QY	260	AATAGACATATTGAACCTACACTCAGATAGTGTTCAGAAATATATCCACCATCGACCAAGTGA	319	
Db	995	TYGGGAGATGYMSYGYCTTGCTTAAAGGAAAGGCMGCTACAGMGAAATYTYGCMGSGYTCKA	1054	

```

RESULT 5
US-10-775-972-153/c
; Sequence 153 Application US/10775972
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaltanya S.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C21
; CURRENT APPLICATION NUMBER: US/10/775,972
; CURRENT FILING DATE: 2004-02-10
; NUMBER OF SEQ ID NOS: 563
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 153
; LENGTH: 2007
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-972-153

```

RESULT 7
US-10-021-698A-796/c
Sequence 796, Application US/10021698A
GENERAL INFORMATION:
APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN EERDEMEGH, PAUL
APPLICANT: DUPUIS, JOSES
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON

PRIOR FILING DATE: 2003-02-28
NUMBER OF SEQ ID NOS: 254
SOFTWARE: PatentIn version 3.2
SEQ ID NO 27
LENGTH: 1575
TYPE: DNA
ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.5;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 317 GGAAGCTAAATAAATCCGATATTACTGCGGCATACGCGGTAATACGCGGCTGG 376
Db 688 GGATTCATATACAGCTTACCGCTGCGATGCTCTGCTGCTGCTGCTGCTGCTGCTG 629
Qy 377 TTAATCAGACCGCATCTGATTCAGCGTAATGTCGCTCAGGTTGG 422
Db 628 TGCAACTGCTGGCTCTCTCTCCAAAGACTTGGGAGCTCAGGTTGG 583

RESULT 11
US-10-767-701-1287/C
Sequence 1287, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53535)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 1287
LENGTH: 883
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
NAME/KEY: unsure
LOCATION: (1) ..(883)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAV03-CLUS19621_1
US-10-767-701-1287

Query Match 6.7%; Score 30.4; DB 6; Length 883;
Best Local Similarity 61.2%; Pred. No. 1.5;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 52 GCTCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111
Db 493 GGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 434
Qy 112 AGTTCGCGCGCGGACTATGA 131
Db 433 CTTTCGCGCGCTTGGGATGA 414

RESULT 12
US-10-767-795-6329/C
Sequence 6329, Application US/10767795
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Cao, Yongwei
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53534)B
CURRENT APPLICATION NUMBER: US/10/767,795
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 117596

SEQ ID NO 6329
LENGTH: 989
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: Clone ID: GOSHI-09MAV01-CL002_1
US-10-767-795-6329

Query Match 6.6%; Score 30.2; DB 6; Length 989;
Best Local Similarity 49.7%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 24 ATTGCGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 83
Db 842 ATTAACGAAACAGACATCTTACATTAATGATCTGCTGCTGCTGCTGCTGCTGCTG 783
Qy 84 CGGCGTAATCATTAACGCGCGCAATAGTTCGCGCGGAGCTATGATGATGATGATG 143
Db 782 TGT 723
Qy 144 CCGT 178
Db 722 CCGTATGAAACAGCAGCTGAGCAACATAC 688

RESULT 13
US-10-767-701-12486/C
Sequence 12486, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53535)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 12486
LENGTH: 1203
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
NAME/KEY: unsure
LOCATION: (1) ..(1203)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAV03-CLUS13462_1
US-10-767-701-12486

Query Match 6.6%; Score 30; DB 6; Length 1203;
Best Local Similarity 48.3%; Pred. No. 2.4;
Matches 84; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 20 CAGATTGCGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 79
Db 271 CCGGAAACCGCTTAAGTGCAGAGTCCCGCGTGGCTTGTGTGTGTGTGTGTGTGTGT 212
Qy 80 GCGCGCGGTAATCATTAACGCGCGCAATGTTCCGCGCGGAGCTATGATGATGATG 139
Db 211 CCGGCGCGGAGTGCAGCGCGGCGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 152
Qy 140 TTACCGGT 193
Db 151 ATGAGCTGCAGTGCAGCGCGCGCTGCGAGCCGAGGAAAGTGTGTGTGTGTGTGTGT 98

RESULT 14
US-10-767-701-6233/C
Sequence 6233, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53535)B

;; CURRENT APPLICATION NUMBER: US/10/767,701
;; CURRENT FILING DATE: 2004-01-29
;; NUMBER OF SEQ ID NOS: 63128
;; SEQ ID NO 6233
;; LENGTH: 1228
;; TYPE: DNA
;; ORGANISM: Sorghum bicolor
;; FEATURE:
;; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS37536_1
US-10-767-701-6233

Query Match 6.5%; Score 29.8; DB 6; Length 1228;
Best Local Similarity 52.9%; Pred. No. 2.8;
Matches 64; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 59 CTGGCCCTCTTCCACATGCGCGCGGATCATACGCGCGGCAATAGTCCG 118
DB 666 CTGCGCGTGAAGCGAGTGAAGCGGATGATGTTCTTGCGGAGGCCCTCGGCGAG 607
QY 119 GCCCGACTATGATGAGCTGTACCGGTGTTTACCATGAATGGACATGCACTGC 178
DB 606 GCGTCATCATGATGAGAGCTCATAGATTCTGTTGCTGAGGCGAGGTGGCTGC 547
QY 179 A 179
DB 546 A 546

RESULT 15
US-10-767-701-20611/C
;; Sequence 20611, Application US/10767701
;; GENERAL INFORMATION:
;; APPLICANT: Kovalic, David K.
;; APPLICANT: Zhou, Yihua
;; APPLICANT: Cao, Yongwei
;; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
;; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
;; FILE REFERENCE: 38-21(5355)B
;; CURRENT APPLICATION NUMBER: US/10/767,701
;; CURRENT FILING DATE: 2004-01-29
;; NUMBER OF SEQ ID NOS: 63128
;; SEQ ID NO 20611
;; LENGTH: 656
;; TYPE: DNA
;; ORGANISM: Sorghum bicolor
;; FEATURE:
;; OTHER INFORMATION: Clone ID: 11064914
US-10-767-701-20611

Query Match 6.5%; Score 29.6; DB 6; Length 656;
Best Local Similarity 52.4%; Pred. No. 2.5;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 28 GCAGCAATCGTAGTTCTGCGAGTCTCTGCTGCGCTGCGCTTCCACATGCGCGCGCGC 87
DB 126 GTAGAAAGTTGCGAGCCTGCGCTTGTGAGTCTGCGGATCATAGGTTGCGGCGTCC 67
QY 88 GGTATCATAAAGCGCGCGGCAATAGTTCGCGCGGACTATGATGAGCTGTACCGCT 147
DB 66 GGTGACCTTGTGACGCGCGGCGAGCGCGGAGCTGACGAGAACCCGACGACAGCT 7
QY 148 GTTG 151
DB 6 GCTG 3

Search completed: March 17, 2004, 08:25:47
Job time: 70.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456
Sequence: 1 atgaacttttaaaagf9gc.....ccacgcttaaccagctatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending Patente NA Main:

```
1: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092a_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092b_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093a_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093b_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095a_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095b_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095c_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095d_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096a_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096b_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096c_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096d_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US097a_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097b_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097c_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097d_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098a_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098b_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098c_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098d_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099a_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099b_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099c_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099d_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099f_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099f_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100a_COMB.seq:*
```

```
44: /cgn2_6/ptodata/2/pna/US100b_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101a_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101b_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102a_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102b_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103a_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103b_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104a_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104b_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107a_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107b_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023a_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023b_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	23	Sequence 19, Appl
2	412.8	90.5	456	6	Sequence 36, Appl
3	412.8	90.5	456	23	Sequence 1, Appl
4	385.6	84.6	456	23	Sequence 25, Appl
5	363.2	79.3	456	23	Sequence 17, Appl
6	361.6	79.3	456	23	Sequence 11, Appl
7	361.6	79.3	456	23	Sequence 13, Appl
8	361.6	79.3	456	23	Sequence 23, Appl
9	360	78.9	456	23	Sequence 15, Appl
10	358.4	78.9	456	23	Sequence 27, Appl
11	358.4	78.9	456	23	Sequence 21, Appl
12	355.2	77.9	456	23	Sequence 29, Appl
13	288.6	63.3	361	6	Sequence 54, Appl
14	236.8	51.9	456	23	Sequence 3, Appl
15	236.8	51.9	456	23	Sequence 3, Appl
16	236.8	51.9	456	23	Sequence 3, Appl
17	234.4	51.4	453	92	Sequence 3, Appl
18	234.4	51.4	453	92	Sequence 1, Appl
19	227.2	49.8	477	17	Sequence 1, Appl
20	227.2	49.8	477	17	Sequence 172, Appl
21	227.2	49.8	477	17	Sequence 172, Appl
22	192.8	42.3	336	13	Sequence 172, Appl
23	168	36.8	360	31	Sequence 1, Appl
24	64.8	14.2	522	17	Sequence 1, Appl
25	64.8	14.2	522	17	Sequence 170, Appl
26	64.8	14.2	522	51	Sequence 170, Appl
27	51.2	11.2	100	45	Sequence 77, Appl
28	50	11.0	78	23	Sequence 52, Appl
29	50	11.0	78	23	Sequence 53, Appl
30	48	10.5	78	23	Sequence 9, Appl
31	45	9.9	78	23	Sequence 49, Appl
32	43.4	9.5	78	23	Sequence 48, Appl
33	42.4	9.3	100	45	Sequence 5066, Ap
34	42	9.2	1290	37	Sequence 5263, Ap
35	36.8	8.1	1290	37	Sequence 5066, Ap
36	36.8	8.1	1290	37	Sequence 23539, A
37	36.2	7.9	632	33	Sequence 61621, A
38	36.2	7.9	805	26	Sequence 61413, A
39	36.2	7.9	805	30	Sequence 61323, A
40	36.2	7.9	805	33	Sequence 61323, A
41	35.8	7.9	387	37	Sequence 217377, Ap
42	35.8	7.9	387	37	Sequence 217377, Ap
43	35.8	7.9	1281	20	Sequence 7264, Ap
44	35	7.7	225664	52	Sequence 1, Appl
45	34.8	7.6	542	24	Sequence 14113, A

ALIGNMENTS

RESULT 1
US-09-543-407-19
Sequence 19, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-19

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 2, 1e-136;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAACTTTAAAGTGGAGCATTCGACGATCGTGTTCGCGAGTGTCTGGCT 60
1 ATGAACTTTAAAGTGGAGCATTCGACGATCGTGTTCGCGAGTGTCTGGCT 60
61 GGGCTGCTTCCAAATGGGGCGGGCGGATCATTAACGGCGCGCAATAGTCCGGC 120
61 GGGCTGCTTCCAAATGGGGCGGGCGGATCATTAACGGCGCGCAATAGTCCGGC 120
121 CCGAGCTCAACGTTGAGCATTTATCAGTACGGTCCGTTAAGCTTCTCTGCGAA 180
121 CCGAGCTCAACGTTGAGCATTTATCAGTACGGTCCGTTAAGCTTCTCTGCGAA 180
181 AGGATGCCCGTAAATATATATAGTACGCTGTTACCGGTGTTTACCAATGACAT 240
181 AGGATGCCCGTAAATATATATAGTACGCTGTTACCGGTGTTTACCAATGACAT 240
241 GCAGGCGAGGCGGATATATATATATATATATATATATATATATATATATAT 300
241 GCAGGCGAGGCGGATATATATATATATATATATATATATATATATATATAT 300
301 GCCACCATGACGACGATGGAAGCTTAAATCTCGATATATATATATATATATAT 360
301 GCCACCATGACGACGATGGAAGCTTAAATCTCGATATATATATATATATATAT 360
361 AATAAGCGCGGCTGTTATATATATATATATATATATATATATATATATATAT 420
361 AATAAGCGCGGCTGTTATATATATATATATATATATATATATATATATATAT 420
421 GGTGTTGCAACAAAGCCGCGCTTACGATTTAA 456
421 GGTGTTGCAACAAAGCCGCGCTTACGATTTAA 456

RESULT 2
US-08-233-642A-56
Sequence 56, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSER: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 90.5%; Score 412.8; DB 6; Length 456;
Best Local Similarity 94.1%; Pred. No. 2.1e-122;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGCAGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGCAGTCTGCT 60
QY 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGGAATAGTTCCGGC 120
DB 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGGAATAGTTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGATACGGTTCGGCTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGATACGGTTCGGCTAACGCTGCTGCTGCA 180
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGATACGGTTCGGCTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGATACGGTTCGGCTAACGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAATATGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
DB 181 AGCGATGCCCGTAATATGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 241 GCAGGCGAGGGGGGGGATATATGATCTATGTAATGATGATGATGATGATGATGAT 300
DB 241 GTAGGCGAGGGGGGGGATATATGATCTATGTAATGATGATGATGATGATGATGAT 300
QY 301 GCCACATCGACAGTGGAAAGCTTAAATCTCCATATTAATGCTGCGCAATACGGCGGT 360
DB 301 GCCACATCGACAGTGGAAAGCTTAAATCTCCATATTAATGCTGCGCAATACGGCGGT 360
QY 361 AATAACGGCGGCTGTTATATCATGACCGCATCTGATTCACGATATGCTGCTGCT 420
DB 361 AATAACGGCGGCTGTTATATCATGACCGCATCTGATTCACGATATGCTGCTGCT 420
QY 421 GGTTCGCAACAACGGCAGCGCTTAACGATTTAA 456
DB 421 GGTTCGCAACAACGGCAGCGCTTAACGATTTAA 456

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1

Query Match 90.5%; Score 412.8; DB 23; Length 456;
Best Local Similarity 94.1%; Pred. No. 2.1e-122;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGCAGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGCAGTCTGCT 60
QY 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGGAATAGTTCCGGC 120
DB 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGGAATAGTTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGATACGGTTCGGCTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGATACGGTTCGGCTAACGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAATATGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
DB 181 AGCGATGCCCGTAATATGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
QY 241 GCAGGCGAGGGGGGGGATATATGATCTATGTAATGATGATGATGATGATGATGAT 300
DB 241 GTAGGCGAGGGGGGGGATATATGATCTATGTAATGATGATGATGATGATGATGAT 300
QY 301 GCCACATCGACAGTGGAAAGCTTAAATCTCCATATTAATGCTGCGCAATACGGCGGT 360
DB 301 GCCACATCGACAGTGGAAAGCTTAAATCTCCATATTAATGCTGCGCAATACGGCGGT 360
QY 361 AATAACGGCGGCTGTTATATCATGACCGCATCTGATTCACGATATGCTGCTGCT 420
DB 361 AATAACGGCGGCTGTTATATCATGACCGCATCTGATTCACGATATGCTGCTGCT 420
QY 421 GGTTCGCAACAACGGCAGCGCTTAACGATTTAA 456
DB 421 GGTTCGCAACAACGGCAGCGCTTAACGATTTAA 456

RESULT 4
US-09-543-407-25
Sequence 25, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding P73 from GP63 of Leishmania major.
US-09-543-407-25

Query Match 84.6%; Score 385.6; DB 23; Length 456;
Best Local Similarity 90.4%; Pred. No. 1.4e-113;
Matches 412; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGCAGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGCAGTCTGCT 60
QY 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGGAATAGTTCCGGC 120

Db 61 GCGCTGTTCCACATGCGGCGCGGTATCATACCGCGCGCAATGTTCCGCG 120
Qy 121 CCGACATCAAGTTGAGCATTTATCAGTACGTTTCCGCTTACCGTCCGCTGCA 180
Db 121 CCGACATCAAGTTGAGCATTTATCAGTACGTTTCCGCTTACCGTCCGCTGCA 180
Qy 181 AGCGATGCGCGTAAATATGATCAGCTGTTACCGGTTGTTTACCATGAATGCGCAT 240
Db 181 CTGGTACCGGTTGTTTACCATGAATGCGCATGAGTTATGTTACCGGCGCAT 240
Qy 241 GCGGCGGAGGTCGAGTAAATAGTACTATTGAAGTACGAGTGTTCAGAAATAT 300
Db 241 GTAGGCGAGGTCGAGTAAATAGTACTATTGAAGTACGAGTGTTCAGAAATAT 300
Qy 301 GCCACATGACAGTGAAGCGTAAATCTCGATATTAATGTTGCGCAATGCGCGGT 360
Db 301 GCCACATGACAGTGAAGCGTAAATCTCGATATTAATGTTGCGCAATGCGCGGT 360
Qy 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Db 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Qy 421 GGTITGGCAACAAGCGCGCTTACCATGATTAA 456
Db 421 GGTITGGCAACAAGCGCGCTTACCATGATTAA 456

RESULT 5

US-09-543-407-17
Sequence 17, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-17

Query Match
Best Local Similarity 79.6%; Score 363.2; DB 23; Length 456;
Matches 198; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGCAGCATTTGCGCATGTTGCTGCGAGTCTGCGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTTGCGCATGTTGCTGCGAGTCTGCGCT 60
Qy 61 GCGCTGTTCCACATGCGGCGCGGTAAATCAATACGCGGCGCAATGTTCCGCG 120
Db 61 GCGCTGTTCCACATGCGGCGCGGTAAATCAATACGCGGCGCAATGTTCCGCG 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACCGTGTCTGCGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACCGTGTCTGCGCA 180
Qy 181 AGGATGCGCGTAAATATGATCAGCTGTTACCGGTTGTTTACCATGAATGCGCAT 240
Db 181 AGGATGCGCGTAAATATGATCAGCTGTTACCGGTTGTTTACCATGAATGCGCAT 240

Qy 241 GCGGCGGAGGTCGAGTAAATAGTACTATTGAAGTACGAGTGTTCAGAAATAT 300
Db 241 GTAGGCGAGGTCGAGTAAATAGTACTATTGAAGTACGAGTGTTCAGAAATAT 300
Qy 301 GCCACATGACAGTGAAGCGTAAATCTCGATATTAATGTTGCGCAATGCGCGGT 360
Db 301 GCCACATGACAGTGAAGCGTAAATCTCGATATTAATGTTGCGCAATGCGCGGT 360
Qy 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Db 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Qy 421 GGTITGGCAACAAGCGCGCTTACCATGATTAA 456
Db 421 GGTITGGCAACAAGCGCGCTTACCATGATTAA 456

RESULT 6

US-09-543-407-11
Sequence 11, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-11

Query Match
Best Local Similarity 79.3%; Score 361.6; DB 23; Length 456;
Matches 197; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGCAGCATTTGCGCATGTTGCTGCGAGTCTGCGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTTGCGCATGTTGCTGCGAGTCTGCGCT 60
Qy 61 GCGCTGTTCCACATGCGGCGCGGTAAATCAATACGCGGCGCAATGTTCCGCG 120
Db 61 GCGCTGTTCCACATGCGGCGCGGTAAATCAATACGCGGCGCAATGTTCCGCG 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACCGTGTCTGCGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACCGTGTCTGCGCA 180
Qy 181 AGGATGCGCGTAAATATGATCAGCTGTTACCGGTTGTTTACCATGAATGCGCAT 240
Db 181 AGGATGCGCGTAAATATGATCAGCTGTTACCGGTTGTTTACCATGAATGCGCAT 240
Qy 241 GCGGCGGAGGTCGAGTAAATAGTACTATTGAAGTACGAGTGTTCAGAAATAT 300
Db 241 GTAGGCGAGGTCGAGTAAATAGTACTATTGAAGTACGAGTGTTCAGAAATAT 300
Qy 301 GCCACATGACAGTGAAGCGTAAATCTCGATATTAATGTTGCGCAATGCGCGGT 360
Db 301 GCCACATGACAGTGAAGCGTAAATCTCGATATTAATGTTGCGCAATGCGCGGT 360
Qy 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Db 361 AATACGCGCGCTGTTATATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420

Oy 421 GATTGGCAACACGCGCTAACAGTATTA 456
Db 421 GCACATGCACACGCGCTAACAGTATTA 456

RESULT 7

US-09-543-407-13
Sequence 13, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-13

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 8.2e-106;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Oy 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATTCGAGTCTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATTCGAGTCTGCTGCT 60
Oy 61 GCGCTGTTCCACATGGGGGCGCGGCGTAACTAAACGGCGGCAATGTTCCGGC 120
Db 61 GCGCTGTTCCACATGGGGGCGCGGCGTAACTAAACGGCGGCAATGTTCCGGC 120
Oy 121 CCGACTCAACGTTGAGCATTTATGATGAGTTCGCTACGCTGCTGCTGCA 180
Db 121 CCGACTCAACGTTGAGCATTTATGATGAGTTCGCTACGCTGCTGCTGCA 180
Oy 181 AGCGATGCGGTAAATATGATCAGCTGCTTACCCGTTGTTACCCATGAATGCGACAT 240
Db 181 AGCGATGCGGTAAATATGATCAGCTGCTTACCCGTTGTTACCCATGAATGCGACAT 240
Oy 241 GCAGCCGAGGTCGGATTAATGATGATTAATGATGATGATGATGATGATGATGAT 300
Db 241 GCAGCCGAGGTCGGATTAATGATGATTAATGATGATGATGATGATGATGATGAT 300
Oy 301 GCGACATCGACGAGTGAAGCTTAAACCTCCGATTAATGATGATGATGATGATGAT 360
Db 301 GCGACATCGACGAGTGAAGCTTAAACCTCCGATTAATGATGATGATGATGATGAT 360
Oy 361 AATAACGCGGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGAGGTT 420
Db 361 CTGTTGCGGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGAGGTT 420
Oy 421 GATTGGCAACACGCGCTAACAGTATTA 456
Db 421 GATTGGCAACACGCGCTAACAGTATTA 456

RESULT 8
US-09-543-407-23
Sequence 23, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-23

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 8.2e-106;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Oy 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATTCGAGTCTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATTCGAGTCTGCTGCT 60
Oy 61 GCGCTGTTCCACATGGGGGCGCGGCGTAACTAAACGGCGGCAATGTTCCGGC 120
Db 61 GCGCTGTTCCACATGGGGGCGCGGCGTAACTAAACGGCGGCAATGTTCCGGC 120
Oy 121 CCGACTCAACGTTGAGCATTTATGATGAGTTCGCTACGCTGCTGCTGCA 180
Db 121 CCGACTCAACGTTGAGCATTTATGATGAGTTCGCTACGCTGCTGCTGCA 180
Oy 181 AGCGATGCGGTAAATATGATCAGCTGCTTACCCGTTGTTACCCATGAATGCGACAT 240
Db 181 AGCGATGCGGTAAATATGATCAGCTGCTTACCCGTTGTTACCCATGAATGCGACAT 240
Oy 241 GCAGCCGAGGTCGGATTAATGATGATTAATGATGATGATGATGATGATGATGAT 300
Db 241 GCAGCCGAGGTCGGATTAATGATGATTAATGATGATGATGATGATGATGATGAT 300
Oy 301 GCGACATCGACGAGTGAAGCTTAAACCTCCGATTAATGATGATGATGATGATGAT 360
Db 301 GCGACATCGACGAGTGAAGCTTAAACCTCCGATTAATGATGATGATGATGATGAT 360
Oy 361 AATAACGCGGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGAGGTT 420
Db 361 CATGAATGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGAGGTT 420
Oy 421 GATTGGCAACACGCGCTAACAGTATTA 456
Db 421 GATTGGCAACACGCGCTAACAGTATTA 456

RESULT 9
US-09-543-407-15
Sequence 15, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15

LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
 US-09-543-407-15

Query Match
 Best Local Similarity 78.9%; Score 360; DB 23; Length 456;
 Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
 Db 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
 QY 61 GGGCTGTTTCCAAATGGGCGGCGGTAAATCATACGGGCGGCAATAGTTCCGGC 120
 Db 61 GGGCTGTTTATGATGATGCTGTTACCCGTTGTTACCCATGAATGACATGCTCCGC 120
 QY 121 CCGGACTCAAGTGTGACATTTATGATGCTGCTTCCGCTTACGCTGCTGCTGCA 180
 Db 121 CCGGACTCAAGTGTGACATTTATGATGCTGCTTCCGCTTACGCTGCTGCTGCA 180
 QY 181 AGCGATGCCGTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
 Db 181 AGCGATGCCGTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
 QY 241 GCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 300
 Db 241 GTRAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 300
 QY 301 GCGACCATGACAGTGAAGCGTAAACCTCCGATATTAATGCTGCTGCGCAATAGCGGCT 360
 Db 301 GCGACCATGACAGTGAAGCGTAAACCTCCGATATTAATGCTGCTGCGCAATAGCGGCT 360
 QY 361 AATAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 420
 Db 361 AATAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 420
 QY 421 GGTITGGCAACGACGCGGCTAACCGATTTAA 456
 Db 421 GGTITGGCAACGACGCGGCTAACCGATTTAA 456

RESULT 10
 US-09-543-407-27

Sequence 27, Application US/09543407
 GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
 FILE REFERENCE: 920043.406
 CURRENT APPLICATION NUMBER: US/09/543,407
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 27
 LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
 US-09-543-407-27

Query Match
 Best Local Similarity 78.9%; Score 360; DB 23; Length 456;
 Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
 Db 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
 QY 61 GGGCTGTTTCCAAATGGGCGGCGGTAAATCATACGGGCGGCAATAGTTCCGGC 120
 Db 61 GGGCTGTTTCCAAATGGGCGGCGGTAAATCATACGGGCGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGTGTGACATTTATGATGCTGCTTCCGCTTACGCTGCTGCTGCA 180
 Db 121 CCGGACTCAAGTGTGACATTTATGATGCTGCTTCCGCTTACGCTGCTGCTGCA 180
 QY 181 AGCGATGCCGTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
 Db 181 AGCGATGCCGTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
 QY 241 GCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 300
 Db 241 TATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
 QY 301 GCGACCATGACAGTGAAGCGTAAACCTCCGATATTAATGCTGCTGCGCAATAGCGGCT 360
 Db 301 GCGACCATGACAGTGAAGCGTAAACCTCCGATATTAATGCTGCTGCGCAATAGCGGCT 360
 QY 361 AATAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 420
 Db 361 AATAAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 420
 QY 421 GGTITGGCAACGACGCGGCTAACCGATTTAA 456
 Db 421 GGTITGGCAACGACGCGGCTAACCGATTTAA 456

RESULT 11
 US-09-543-407-21

Sequence 21, Application US/09543407
 GENERAL INFORMATION:
 APPLICANT: White, Aaron P.
 APPLICANT: Doran, James L.
 APPLICANT: Collinson, S. Karen
 TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
 FILE REFERENCE: 920043.406
 CURRENT APPLICATION NUMBER: US/09/543,407
 NUMBER OF SEQ ID NOS: 59
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 21
 LENGTH: 456
 TYPE: DNA
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
 OTHER INFORMATION: sequence containing the replacement fragment
 OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
 US-09-543-407-21

Query Match
 Best Local Similarity 78.6%; Score 358.4; DB 23; Length 456;
 Matches 395; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
 Db 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
 QY 61 GGGCTGTTTCCAAATGGGCGGCGGTAAATCATACGGGCGGCAATAGTTCCGGC 120
 Db 61 GGGCTGTTTCCAAATGGGCGGCGGTAAATCATACGGGCGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGTGTGACATTTATGATGCTGCTTCCGCTTACGCTGCTGCTGCA 180

Db	121	CCGAGCTCAACGTTAGCATTTATACGTAAGGTTCCGCTAACCGCTGCTTGCCTGCA	180
Qy	181	AGCGATGCCCGTAATATGATATAGCTGCTTTACCCGTTGTTTACCATGAAATGGCAAT	240
Db	181	AGCGATGCCCGTAATATCTAATAACGACCATTTACCCAGAGCGGTTATGTAACGGCGCGAT	240
Qy	241	GCAGGCCAGGGGCGGATTAAGTACTATTGAACTGACTCAGAAATGTTTCAAAATTAAT	300
Db	241	GTAGGCCAGGGGCGGATTAATTATGATCAGCTGTTACCCGTTGTTTACCATGAATG	300
Qy	301	GCCACCATGACACAGTGAACGCTAAATACTCCGATATTACTTCCGCCATACGGCGGT	360
Db	301	GCACATGACAGACAGTGGAAACGCTAATAAACTCCGATATTACTTCCGCCAATACGGCGGT	360
Qy	361	AATAACGGCGCGCTGTTATCAGACCGCATCTGATTTCCAGCGTAATGCTGCGTCAAGTT	420
Db	361	AATAACGGCGCGCTGTTAATCAGACCGCATCTGATTTCCAGCGTAATGCTGCGTCAAGTT	420
Qy	421	GGTTTGGCAACAGCGCCAGGCTAACCAAGTATTA	456
Db	421	GGTTTGGCAACAGCGCCAGGCTAACCAAGTATTA	456

RESULT 12

```

Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Dorian, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLARY SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

```

Query Match	77.9%	Score 355.2	DB 23	Length 456
-------------	-------	-------------	-------	------------

```
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

Qy	1	ATGAAACTTTTAAAGTGCAGACATTCGAGCAATGTAAGTTCTGCGAGTGTCTGGCT	60
Db	1	ATGAAACTTTTAAAGTGCAGACATTCGAGCAATGTAAGTTCTGCGAGTGTCTGGCT	60
Qy	61	GAGCTGTTCCAAATGGGGCGGGCGGGTAATCATTAACGGCGCGGCAATAGTTCCGGC	120
Db	61	GAGCTGTTCCCAATGGGGCGGGCGGGTAATCATTAACGGCGCGGCAATAGTTCCGGC	120
Qy	121	CCGAGCTAAAGCTTAGACATTATACATGACGATCGGTCGCTAACGCTGCTGTCGCA	180
Db	121	CCGAGCTAAAGCTTAGACATTATACATGACGATCGGTCGCTAACGCTGCTGTCGCA	180
Qy	181	AGCGATGCCCGTAAATATATGATCAGCTGTTACCCCGTGTGTTACCCATGAATATGACAT	240
Db	181	AGCGATGCCCGTAAATATGTAACGACCAATTACCCAGAGCGGTTATGTTAATGCGCGCGAT	240
Qy	241	GCAGGCCAGGGCGGATTAATAGTACTATTGAACGTACACAGAAATGCTTTCAGAAATTAAT	300
Db	241	GTAGGCCAGGGCGGATTAATAGTACTATTGAACGTACACAGAAATGCTTTCAGAAATTAAT	300

Qy	361	GCACCAATGCACCACTGGAACGGCTAAAACTCCGATATATACGTGGGCCAATTACGGCGGT	366
Db	301	GCCACTATGATGATCACTGCTGTTACCCGTCGTTGTTATCCAGTAATGCAATGCACATGCAGCGCGT	366
Qy	361	AATAACGCCGCGCGCTGGTTAATCAGACCGCATCTGATTCAGCGGTAAATGGTGCGCTCAGGTT	420
Db	361	AATAACGCCGCGCGCTGGTTAATCAGACCGCATCTGATTCAGCGGTAAATGGTGCGCTCAGGTT	420
Qy	421	GGTTTGGCAACAACGCCACGGCTTAACCAAGTATTAA	456
Db	421	GGTTTGGCAACAACGCCACGGCTTAACCAAGTATTAA	456

RESULT 13

```

1 Sequence 64, Application US/08233642A
2 GENERAL INFORMATION:
3 APPLICANT: Kay, William W.
4 APPLICANT: Collinson, S. Karen
5 APPLICANT: Cloutier, Sharon C.
6 APPLICANT: Doran, James L.
7 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA
8 TITLE OF INVENTION: BASED VACCINES
9 NUMBER OF SEQUENCES: 58
10 NUMBER OF SEQUENCES: -
11 CORRESPONDENCE ADDRESS:
12 ADDRESSEE: Seed and Berry
13 STREET: 6300 Columbia Center, 701 Fifth Avenue
14 CITY: Seattle
15 STATE: Washington
16 COUNTRY: U.S.A.
17 ZIP: 98104-7092
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: Floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: Patent In Release #1.0, Version #1.25
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/233,642A
25 FILING DATE: 26-APR-1994
26 CLASSIFICATION: 424
27 ATTORNEY/AGENT INFORMATION:
28 NAME: King, Joshua
29 REGISTRATION NUMBER: 35,570
30 REFERENCE/DOCKET NUMBER: 920043, 403C3
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (206) 622-4900
33 TELEFAX: (206) 682-6031
34 TELEK: 3723836 SEEDANBERRY
35 INFORMATION FOR SEQ ID NO: 54:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 361 base pairs
38 TYPE: nucleic acid
39 STRANDEDNESS: double
40 TOPOLOGY: linear
41 MOLECULE TYPE: DNA (genomic)
42 FEATURE:
43 NAME/KEY: CDS
44 LOCATION: 1..357
45 US-08-233-642A-54

```

Query Match	63.3%;	Score 288.6;	DB 6;	Length 361;
-------------	--------	--------------	-------	-------------

Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Oy	64	GTCTGTCACACATGGGGCGGCGCGGTATCATTAACGGCGCGGCAATGTTCCGGCCG	123
Db	1	GTCTGTCACACATGGGGCGGCGCGGTATCATTAACGGCGCGGCAATGTTCCGGCCG	60
Oy	124	GACTCAAGTGTGAGATTATCAGTACGGTTCCGCTAACGCTGGCTTCTCTGCACAC	183
Db	61	GACTCAAGTGTGAGATTATCAGTACGGTTCCGCTAACGCTGGCTTCTCTGCACAC	120

```

QY 184 GATCCCGGTAATATGATAGCTGTTACCCGTTGTTACCATGAATGCAATGCA 243
DB 121 GATGCCCGTAATCTGAAACGACATTAACCCAGACGGTTATGTAAGCGGCGGATGTA 180
QY 244 GGCACAGGTTGCGGATATATGACTATTTGAAGTCACTGCAATGTTTCAAAATATATGCC 303
DB 181 GGCACAGGTTGCGGATATATGACTATTTGAAGTCACTGCAATGTTTCAAAATATATGCC 240
QY 304 ACCATGACCACTGGAAGCGCTTAAATCTCGCATATTTACTGTGCGCAATACGGCGGTAAT 363
DB 241 ACCATGACCACTGGAAGCGCTTAAATCTCGCATATTTACTGTGCGCAATACGGCGGTAAT 300
QY 364 AACGCGCGCTGTTATATCAGACCGCATCTGATTC 398
DB 301 AACGCGCGCTGTTATATCAGACCGCATCTGATTC 335

```

```

RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli
US-09-543-407-3

```

```

Query Match 51.9%; Score 236.8; DB 23; Length 456;
Best Local Similarity 70.0%; Pred. No. 2.3e-65;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGAGTTTGTGCACTGCTGGCT 60
DB 1 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGAGTTTGTGCACTGCTGGCT 60
QY 61 GCGCTGTTCCAAATGCGGCGCGGCTATCATACGCGCGGCAATAGTTCCGCG 120
DB 61 GCGCTGTTCCAAATGCGGCGCGGCTATCATACGCGCGGCAATAGTTCCGCG 120
QY 121 CCGAATTCGAGTGAACATTTACGAGTGGCGGTAACCGGTAATATATACGCG 120
DB 121 CCGAATTCGAGTGAACATTTACGAGTGGCGGTAACCGGTAATATATACGCG 120
QY 181 AGGATGCCCGTAATATGATCACTGCTGTTACCCGTTGTTACCCATGAATGACAT 240
DB 181 AGGATGCCCGTAATATGATCACTGCTGTTACCCGTTGTTACCCATGAATGACAT 240
QY 241 GCAAGCCAGGTTGCGGATATATGATCACTGCTGTTACCCGTTGTTACCCATGAATGACAT 240
DB 241 GCAAGCCAGGTTGCGGATATATGATCACTGCTGTTACCCGTTGTTACCCATGAATGACAT 240
QY 301 GCGACCATGACCACTGGAAGCGCTTAAATCTCGCATATTTACTGTGCGCAATACGGCGGT 360
DB 301 GCGACCATGACCACTGGAAGCGCTTAAATCTCGCATATTTACTGTGCGCAATACGGCGGT 360
QY 361 AATTAAGCGCGGCTGTTATATCAGACCGCATCTGATTTCAAGGTAATGTCGTCAGATT 420
DB 361 AATTAAGCGCGGCTGTTATATCAGACCGCATCTGATTTCAAGGTAATGTCGTCAGATT 420
QY 421 GGTGTTGCAACCAAGCGGCTATACCAAGATTAA 456
DB 421 GGTGTTGCAACCAAGCGGCTATACCAAGATTAA 456

```

```

RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: OLSEN, Arne
; APPLICANT: NORMARK, Staffan
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978.878
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)...(538)
US-08-978-878-3

```

```

Query Match 51.9%; Score 236.8; DB 13; Length 675;
Best Local Similarity 70.0%; Pred. No. 2.6e-65;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGAGTTTGTGCACTGCTGGCT 60
DB 83 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGAGTTTGTGCACTGCTGGCT 60
QY 61 GCGCTGTTCCAAATGCGGCGCGGCTATCATACGCGCGGCAATAGTTCCGCG 120
DB 143 GCGCTGTTCCAAATGCGGCGCGGCTATCATACGCGCGGCAATAGTTCCGCG 120
QY 121 CCGAATTCGAGTGAACATTTACGAGTGGCGGTAACCGGTAATATATACGCG 202
DB 121 CCGAATTCGAGTGAACATTTACGAGTGGCGGTAACCGGTAATATATACGCG 202
QY 181 AGGATGCCCGTAATATGATCACTGCTGTTACCCGTTGTTACCCATGAATGACAT 240
DB 203 CCAATTCGAGTGAACATTTACGAGTGGCGGTAACCGGTAATATATACGCG 262
QY 241 GCAAGCCAGGTTGCGGATATATGATCACTGCTGTTACCCGTTGTTACCCATGAATGACAT 240
DB 241 GCAAGCCAGGTTGCGGATATATGATCACTGCTGTTACCCGTTGTTACCCATGAATGACAT 240
QY 301 GCGACCATGACCACTGGAAGCGCTTAAATCTCGCATATTTACTGTGCGCAATACGGCGGT 360
DB 383 GCGACCATGACCACTGGAAGCGCTTAAATCTCGCATATTTACTGTGCGCAATACGGCGGT 360
QY 361 AATTAAGCGCGGCTGTTATATCAGACCGCATCTGATTTCAAGGTAATGTCGTCAGATT 420
DB 443 GCGAAGCGGCTGTTATATCAGACCGCATCTGATTTCAAGGTAATGTCGTCAGATT 420
QY 421 GGTGTTGCAACCAAGCGGCTATACCAAGATTAA 456
DB 503 GGTGTTGCAACCAAGCGGCTATACCAAGATTAA 538

```

Thu Mar 18 12:28:03 2004

us-09-543-407-19.rmpm

Page 9

Search completed: March 18, 2004, 02:47:09
Job time : 2925.24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaacttcaaaagtgsc.....ccagcgctacacgtrataa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA New :
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236.8	51.9	675	US-09-741-873C-3	Sequence 3, Appli
2	192.8	42.3	396	US-09-741-873C-1	Sequence 1, Appli
3	30.8	6.8	1575	US-10-788-792-27	Sequence 27, Appli
4	30.4	6.7	848	US-10-100-683-643	Sequence 643, App
5	29.4	6.4	240	US-10-767-701-28437	Sequence 28437, A
6	29.2	6.4	249	US-10-771-241-117	Sequence 117, App
7	29.2	6.4	1046	US-10-771-241-29	Sequence 29, Appli
8	29	6.4	1046	US-10-048-770C-3	Sequence 3, Appli
9	28.8	6.3	83009	PCT-US04-04280-78	Sequence 78, Appli
10	28.8	6.3	1186	US-10-417-375A-143	Sequence 143, App
11	28.6	6.3	1295	PCT-US04-05654-1030	Sequence 1030, App
12	28.6	6.3	3416	US-10-100-683-5480	Sequence 5480, Ap
13	28.6	6.3	8206	US-10-100-683-3397	Sequence 3397, Ap
14	28.4	6.2	394468	US-10-021-698A-924	Sequence 3529, Ap
15	28.4	6.2	201	US-60-548-091-5725	Sequence 5725, Ap
16	28.2	6.2	596	US-60-548-091-32688	Sequence 32688, A
17	28.2	6.2	610	US-10-767-701-5378	Sequence 5378, Ap
18	28.2	6.2	1746	US-10-767-701-8781	Sequence 8781, Ap
19	28.2	6.2	14449	PCT-US04-05654-2555	Sequence 2555, Ap
20	28.2	6.2	225724	US-10-021-698A-924	Sequence 924, App
21	28.2	6.1	237	US-10-021-698A-711	Sequence 711, App
22	28	6.1	554	US-10-767-701-31417	Sequence 31417, A
23	28	6.1	749	US-10-767-701-37293	Sequence 37293, A
24	27.8	6.1	1400	US-10-767-701-12516	Sequence 12516, A
25	27.8	6.1	1400	US-60-545-213-2134	Sequence 2134, Ap
26	27.8	6.1	1400	US-60-545-213-2135	Sequence 2135, Ap

C 27	27.8	6.1	1400	7	US-60-545-213-6406	Sequence 6406, Ap
C 28	27.8	6.1	1400	7	US-60-545-213-6407	Sequence 6407, Ap
C 29	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
C 30	27.6	6.1	363	6	US-10-767-701-29572	Sequence 29572, A
C 31	27.6	6.1	463	6	US-10-767-701-17923	Sequence 17923, A
C 32	27.6	6.1	529	6	US-10-767-701-19620	Sequence 19620, A
C 33	27.6	6.1	551	6	US-10-767-701-3483	Sequence 3483, Ap
C 34	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
C 35	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
C 36	27.6	6.1	617	6	US-10-781-469-57	Sequence 57, Appli
C 37	27.6	6.1	632	6	US-10-100-683-3854	Sequence 3854, Ap
C 38	27.6	6.1	946	6	US-10-784-004-232	Sequence 232, App
C 39	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
C 40	27.6	6.1	1818	6	US-10-773-236-118	Sequence 118, App
C 41	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
C 42	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
C 43	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
C 44	27.6	6.1	8654	6	US-10-453-372-1187	Sequence 1187, Ap
C 45	27.6	6.1	25426	6	US-10-100-683-11419	Sequence 11419, A

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
APPLICANT: Oleen, Arne
TITLE OF INVENTION: Ribonectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-064
CURRENT FILING DATE: US/09/741, 873C
PRIOR APPLICATION NUMBER: 2000-12-22
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978, 878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347, 189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789, 437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970, 846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187, 865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318, 519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 51.9%; Score 236.8; DB 5; Length 675;
Best Local Similarity 70.0%; Pred. No. 3.2e-72;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
OY 1 ATGAACTTTAAAGTGCGAGCATTCGAGCATCGTACTTTCGCGAGTCTGGCT 60
DB 83 ATGAACTTTAAAGTGCGAGCATTCGAGCATCGTACTTTCGCGAGTCTGGCT 142
OY 61 GGGGTGTTCCAAATGCGCGCGCGGTAATCATTAACGCGCGCATATGTCGCGC 120
DB 143 GGGGTGTTCTCAAGTACGCGCGCGGTAACCAACGCGTGTGCGGTAAATAGCGGC 202
OY 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTGCGTAAACGCTGCTTGCACAA 180

Db 203 CCAATTCGTAGCTGAACATTACAGTACGAGTGGCGGTAATCTCTGCACTTGTCTGCAA 262
 Qy 181 AGCGATGCGCGTAAATATGATCAGCTGGTTACCGGTGTTTATCCCATGAATAGCAACAT 240
 Db 263 ACTATATCCCGTAACTCTGACTTGAATATACCAGATGGCGCGGTAATGGTGCAGAT 322
 Qy 241 GCAAGCGAGGCTGGATATATGATCATGATCACTGATCAATGATGTTTCAAGAAATAT 300
 Db 323 GTTGTGTAAGGGCTCAGATGACAGCTCAATGATCTTACCAACGTTGGTTCGTAAACAG 382
 Qy 301 GCCACATGACCAAGTGAAGCGTAAATACTCCGATATATCTGCGCCATATACGGCGGT 360
 Db 383 GCTACTCTTATGATGAGTGAACGGCAAAATTTGAAATAGCGTTAAACAGTTCCGTGT 442
 Qy 361 AATAAGCGCGCGGTGTTATGATCAGACCGATCTGATTCAGCGTAAATGGTCCGTAGATT 420
 Db 443 GCGAACGGTCTCAGATGTTACAGACTGATCTTAACCTCTCCGTCAAGTACTAGATT 502
 Qy 421 GGTTTGGCAACAGCGCCAGCGCTAACAGATTA 456
 Db 503 GGCTTGTGTAACAGCGCACCGCTCATCTAGTACTAA 538

RESULT 2

US-09-741-873C-1
 ; Sequence 1, Application US/09741873C
 ; GENERAL INFORMATION:
 ; APPLICANT: Normark, Staffan
 ; APPLICANT: Olsen, Arne
 ; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
 ; FILE REFERENCE: 012889-084
 ; CURRENT APPLICATION NUMBER: US/09/741, 873C
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: SE 8801723-1
 ; PRIOR FILING DATE: 1998-05-06
 ; PRIOR APPLICATION NUMBER: US 08/976, 878
 ; PRIOR FILING DATE: 1997-11-26
 ; PRIOR APPLICATION NUMBER: US 07/347, 189
 ; PRIOR FILING DATE: 1989-05-04
 ; PRIOR APPLICATION NUMBER: US 07/789, 437
 ; PRIOR FILING DATE: 1991-11-06
 ; PRIOR APPLICATION NUMBER: US 07/970, 846
 ; PRIOR FILING DATE: 1992-11-03
 ; PRIOR APPLICATION NUMBER: US 08/187, 865
 ; PRIOR FILING DATE: 1994-01-28
 ; PRIOR APPLICATION NUMBER: US 08/318, 519
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 396
 ; TYPE: DNA
 ; ORGANISM: Escherichia coli
 ; US-09-741-873C-1

Query Match

Best Local Similarity 42.3%; Score 192.8; DB 5; Length 396;
 Matches 269; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 61 GGGGTGTTTCCAAATGAGGCGCGCGGTATCATTAACGCGCGCAATAGTCCGCG 120
 Db 1 GGTGTGTTTCTTAGTACGCGCGCGCGGTAAACAGGTGTGGGTAAATATAGCGCG 60
 Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTGTTCCGTTCCGTTCCGTTCTGCA 180
 Db 61 CCAATTTGAGCTGACATTTTACAGTACGCTGCGGTGTTGCACTGCTCTGCAA 120
 Qy 181 AGCGATGCGCGTAAATATGATCAGCTGGTTACCGGTGTTTATCCCATGAATAGCA 240
 Db 121 ACTGATGCGCGTAACTGATCTGATTAATCCAGCATGGCGCGGTAAATGGTGCAG 180
 Qy 241 GCAAGCGAGGCTGGATATATGATCATGATCACTGATCAATGATGTTTCAAGAAAT 300
 Db 121 GCGGCGAGGCTGGATATATGATCATGATCACTGATCAATGATGTTTCAAGAAAT 300

Db 181 GTTGTGAGGGCTCAGATGACAGCTCAATGATCTGATCCCAACGTTGGTTACAGC 240
 Qy 301 GCCACATGACCAAGTGAAGCGTAAATACTCCGATATATCTGCGCCATATACGGCGGT 360
 Db 241 GCTACTCTTATGATGAGTGAACGGCAAAATTTGAAATAGCGTTAAACAGTTCCGTGT 300
 Qy 361 AATAAGCGCGCGGTGTTATGATCAGACCGATCTGATTCAGCGTAAATGGTCCGTAGATT 420
 Db 301 GCGAACGGTCTCAGATGTTACAGACTGATCTTAACCTCTCCGTCAAGTACTAGATT 360
 Qy 421 GGTTTGGCAACAGCGCGCTAACAGATTA 456
 Db 361 GGCTTGTGTAACAGCGCACCGCTCATCTAGTACTAA 396

RESULT 3

US-10-788-792-27/c
 ; Sequence 27, Application US/10788792
 ; GENERAL INFORMATION:
 ; APPLICANT: Bayer Pharmaceuticals Corporation
 ; APPLICANT: Eysenck, Deepa
 ; APPLICANT: Bigwood, Douglas
 ; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
 ; FILE REFERENCE: 5152
 ; CURRENT APPLICATION NUMBER: US/10/788, 792
 ; PRIOR FILING DATE: 2004-02-27
 ; PRIOR APPLICATION NUMBER: US 60/450, 655
 ; PRIOR FILING DATE: 2003-02-28
 ; NUMBER OF SEQ ID NOS: 254
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 27
 ; LENGTH: 1575
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-10-788-792-27

Query Match

Best Local Similarity 6.8%; Score 30.8; DB 6; Length 1575;
 Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 317 GGAAGCTTAAATCCGATTTACTGTGCGGCAATAGCGCGGTAAATAGCGCGGTGG 376
 Db 688 GGAATTCATACATGCTGACCGTCACTTGTCTGATTCGTTCAAGAGAGCTGCGCTGG 629
 Qy 377 TTAATCAAGCGCATCTGATTCAGCGGTAAATGGTGCCTAGGTTGG 422
 Db 628 TGCACTGCTGCTCTTCTTCCAAAGATTGGGACTCCAGTTGG 583

RESULT 4

US-10-100-683-643
 ; Sequence 643, Application US/10100683
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen, et al.
 ; TITLE OF INVENTION: Human Secreted Proteins
 ; FILE REFERENCE: PS900
 ; CURRENT APPLICATION NUMBER: US/10/100, 683
 ; PRIOR FILING DATE: 2002-03-19
 ; PRIOR APPLICATION NUMBER: US 60/040, 162
 ; PRIOR FILING DATE: 1997-03-07
 ; PRIOR APPLICATION NUMBER: US 60/043, 576
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047, 601
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056, 845
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: US 60/043, 580
 ; PRIOR FILING DATE: 1997-04-11
 ; PRIOR APPLICATION NUMBER: US 60/047, 599
 ; PRIOR FILING DATE: 1997-05-23
 ; PRIOR APPLICATION NUMBER: US 60/056, 664
 ; PRIOR FILING DATE: 1997-08-22
 ; PRIOR APPLICATION NUMBER: US 60/043, 314

```

RESULT 6
US-10-771-241-117/C
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Forsyth, R. Allyn
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI

```

Qy 299 AATATATGCAACATTCGACCATGTGAAAGCGTAAAACTCCGATATTTACTGTGCGCCCAATG 354

Db 264 AAATTTGAAGCATCGACTTTACCAAGGCAACACAGGTCTGATCTCATCCGCGACATAC 323

Qy 355 GCGCGTATTAACGCGCGCGCTGTTAAACAACCGCATCTGATTTCCAGCGTAATGATGCT 414

Db 324 GGCACCGCATTTCCCGACAGGATAGTCATTAAGAAATTCACCTCCATCTTTACTCGGACT 383

Qy 415 CA 416

DB 384 GA 385

RESULT 8

US-10-048-770C-3/C
 / Sequence 3, Application US/10048770C
 / GENERAL INFORMATION:
 / APPLICANT: NEW BIOTECHNIC S.A.
 / APPLICANT: UNIVERSIDAD DE SEVILLA
 / TITLE OF INVENTION: ANTI-FUNGAL COMPOSITIONS AND METHOD FOR CONTROLLING FUNGI
 / FILE REFERENCE: 1837-2
 / CURRENT APPLICATION NUMBER: US/10/048,770C
 / PRIOR FILING DATE: 2002-01-31
 / PRIOR APPLICATION NUMBER: PCT/ES00/00292
 / NUMBER OF SEQ ID NOS: 28
 / SOFTWARE: PatentIn version 3.1
 / SEQ ID NO 3
 / LENGTH: 1046
 / TYPE: DNA
 / ORGANISM: Trichoderma harzianum
 / US-10-048-770C-3

Query Match
 Best Local Similarity 6.4%; Score 29; DB 6; Length 1046;
 Matches 77; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 157 GCTAACGCTGCGCTGCTCTGCAAGCGATGCGCCGTAATATGATGAGCTGGTAAACCGCT 216
 DB 760 GCAAGCGTGTGATATTCATCATATGAGGTTCGCTAAAGACGGTCTGACCGTTACGAGT 701
 QY 217 GTTGTATCCATTAATGCGACATGCGCGCGGTGCGATATATGATATGAACTG 276
 DB 700 CAGAGCAACTTTTGTCTGCTGCCATGCTGCGCGGATTTGAAAGATGTCCTCCGCGCTT 641
 QY 277 ACTCAAGATGTTTCAGAAATATATGCGACCATGAC 313
 DB 640 GGCCTGGAACGTTTGAAGTGTGCCACGACGCTGACC 604

RESULT 9

PCT-US04-04280-78/C
 / Sequence 78, Application PC/TUS0404280
 / GENERAL INFORMATION:
 / APPLICANT: INCYTE CORPORATION; KABLE, Amy E.;
 / APPLICANT: YUE, Henry; BAUGHN, Mariah R.;
 / APPLICANT: TRIBOULEY, Catherine M.; RING, Huijun Z.;
 / APPLICANT: TRAN, Uyen K.; EBERLING, Brooke E.;
 / APPLICANT: RAMAKRAN, Anita; LEE, Soo Yeun;
 / APPLICANT: SWARNAKAR, Narinder K.; GIETZEN, Kimberly J.;
 / APPLICANT: MARQUIS, Joseph P.; ELLIOTT, Vicki S.;
 / APPLICANT: BECHA, Shanya D.; FAVERO, Kristin D.;
 / APPLICANT: WANG, Jonathan T.; NAIDU, Sangeeta;
 / APPLICANT: HAWKINS, Phillip R.; JIN, Pei;
 / TITLE OF INVENTION: ENZYMES
 / FILE REFERENCE: PF-1507 PCT
 / CURRENT APPLICATION NUMBER: PCT/US04/04280
 / PRIOR FILING DATE: 2004-02-12
 / PRIOR APPLICATION NUMBER: US 60/447,246
 / PRIOR FILING DATE: 2003-02-12
 / PRIOR APPLICATION NUMBER: US 60/449,087
 / PRIOR FILING DATE: 2003-02-21
 / PRIOR APPLICATION NUMBER: US 60/450,622
 / PRIOR FILING DATE: 2003-02-26
 / PRIOR APPLICATION NUMBER: US 60/456,704
 / PRIOR FILING DATE: 2003-03-21
 / PRIOR APPLICATION NUMBER: US 60/463,194
 / PRIOR FILING DATE: 2003-04-15
 / PRIOR APPLICATION NUMBER: US 60/469,358
 / PRIOR FILING DATE: 2003-05-09
 / PRIOR APPLICATION NUMBER: US 60/475,532

/ PRIOR FILING DATE: 2003-06-02
 / PRIOR APPLICATION NUMBER: US 60/476,278
 / PRIOR FILING DATE: 2003-06-04
 / PRIOR APPLICATION NUMBER: US 60/483,395
 / PRIOR FILING DATE: 2003-06-27
 / NUMBER OF SEQ ID NOS: 78
 / SOFTWARE: PERL Program
 / SEQ ID NO 78
 / LENGTH: 3816
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / FEATURE:
 / NAME/KEY: misc_feature
 / OTHER INFORMATION: Incyte ID No: 7517280CB1
 / PCT-US04-04280-78

Query Match
 Best Local Similarity 6.3%; Score 28.8; DB 1; Length 3816;
 Matches 69; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 118 GCGCCGACTCAACGTTGACATTATCATGACGCTTCCGTAACGCTGCTGCTG 177
 DB 2245 GGCATGACAAAGATATGAGAAAGACATGCTGACCTTCTCATGCTTGAAGCTG 2186
 QY 178 CAAAGCATGCGCGTAATATGATGACGCTGTTACCGGTGTTTACCATGAATGCA 237
 DB 2185 CACAGTGTGCTTCCGTAACATGACTCATGACATCATGCTGTGCTGCGCGGACG 2126
 QY 238 CATGAGCGCCAGGCTG 253
 DB 2125 GTGCCAGGCCACTGTG 2110

RESULT 10

US-10-417-375A-143/C
 / Sequence 143, Application US/10417375A
 / GENERAL INFORMATION:
 / APPLICANT: David W. Morris
 / APPLICANT: Marc Malandro
 / TITLE OF INVENTION: Novel Therapeutic Targets in Cancer
 / FILE REFERENCE: 529452001600
 / CURRENT APPLICATION NUMBER: US/10/417,375A
 / NUMBER OF SEQ ID NOS: 176
 / SOFTWARE: RastSeq for Windows Version 4.0
 / SEQ ID NO 143
 / LENGTH: 83009
 / TYPE: DNA
 / ORGANISM: Mus musculus
 / FEATURE:
 / NAME/KEY: misc_feature
 / LOCATION: (1)...(83009)
 / OTHER INFORMATION: n = A,T,C or G
 / US-10-417-375A-143

Query Match
 Best Local Similarity 6.3%; Score 28.8; DB 6; Length 83009;
 Matches 78; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 281 AGAATGTTTCAGAAATATGACCAATGACCAAGTGAAGCGTAATAAAGCTCCGATTTA 340
 DB 14056 AGAATATTAAGAAATGATATATTCCTCAAAAAGAACTTGAATATCTTATTC 13997
 QY 341 CTGTGCGCCAAATAGCGCGGTATTAACCGCGCTGTTAATCAAGCCGATCTGATTCGA 400
 DB 13996 CTTCAAGTACCTTAAGTATGCTTCAATGAATCAAGGACGACCTATTAATTCA 13937
 QY 401 GCGTAATGTCGCTGAGTGTGTTTGGCAACAAGCCAC 440
 DB 13936 GAATATGACACTCAGATGATTAACAGGACGATGCCAC 13897

RESULT 11

PCT-US04-05654-1030
; Sequence 1030, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Haake, Jacqueline B
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omeira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guttererson, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kunitomo, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1030
; LENGTH: 1186
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Predicted polypeptide sequence is orthologous to C652
PCT-US04-05654-1030

Query Match 6.3%; Score 28.6; DB 1; Length 1186;
Best Local Similarity 59.0%; Pred. No. 5.7;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 44 CTGCGAGTCTCTGCTGGCGCTGCTCCACATATGGCGCGCGGCTATATACGCG 103
DB 606 CAGGAGCTGCTCCAGCGCGCGCGCTACGCGGTGCGCGCGGTGTGTGCGCG 665
QY 104 GCGGCAATAGTTCGCGCGCGGAC 126
DB 666 GCTGCTACAACTGCGCGCGGAC 688

RESULT 12
US-10-100-683-5480/c
; Sequence 5480, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599

PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5480
; LENGTH: 1295
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-100-683-5480

Query Match 6.3%; Score 28.6; DB 6; Length 1295;
Best Local Similarity 49.7%; Pred. No. 5.9;
Matches 73; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 242 CAGGCGAGGTGCGGATATAGTACTATTGACTGACTCAGATGTTTCAATATATG 301
DB 767 CGGACCTGCTGCCCCCTAACCTCAAGTCACTGCTGACGCTGCCAGGCGATTAATG 708
QY 302 CCACCATGACCACTGAGCAAGCTTAAATCTCCGATTAATGTCGCCAATACGCGGTA 361
DB 707 ACAGCATCAATCAAGCTCATCATATGTCACCCAGGACCGCGGCAAGAGATGTG 648
QY 362 ATTAAGCGCGCGCTGCTTAATCAGACCG 388
DB 647 ATTAAGCGCTGCGGGAATTGAGACGG 621

RESULT 13
US-10-100-683-3397/c
; Sequence 3397, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3397
; LENGTH: 3416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-100-683-3397

